Energy Efficiency Behavioral Programs:

Literature Review, Benchmarking Analysis, and Evaluation Guidelines

Conservation Applied Research & Development (CARD)
FINAL REPORT

Prepared for: Minnesota Department of Commerce,
Division of Energy Resources

Prepared by: Illume Advising with subcontractors Indicia Consulting
and Dr. Edward Vine

COMM-20140512-86775 | May 4, 2015

Prepared by:

Anne Dougherty, Founder, ILLUME Advising, LLC
Dr. Courtney Henderson, Senior Evaluation Advisor, ILLUME Advising, LLC
Amanda Dwelley, Directing Advisor, ILLUME Advising, LLC
Mallika Jayaraman, Research Coordinator, ILLUME Advising, LLC

Sub-Contractors:
Dr. Edward Vine, Staff Scientist, Lawrence Berkeley National Laboratory
Dr. Susan Mazur-Stommen, Founder, Indica Consulting

440 Science Drive #202
Madison, WI 53711
608-512-3930
www.illumeadvising.com
Anne Dougherty

Contract Number: 86775

Prepared for Minnesota Department of Commerce, Division of Energy Resources

Mike Rothman, Commissioner, Department of Commerce

Bill Grant, Deputy Commissioner, Department of Commerce, Division of Energy Resources

Joe Plummer, Project Manager
651.539.1874
joe.plummer@state.mn.us

ACKNOWLEDGEMENTS

This project was supported in part (or in whole) by a grant from the Minnesota Department of Commerce, Division of Energy Resources, through the Conservation Applied Research and Development (CARD) program, which is funded by Minnesota ratepayers.

The authors would also like to acknowledge the following for their contributions to the project: Joseph Plummer and Jessica Burdette, both from the Minnesota Department of Commerce, Division of Energy Resources.

DISCLAIMER

Table of Contents

This report is divided into two primary sections: (1) the literature review and benchmarking analysis, and (2) evaluation guidelines.

Literature Review, Benchmarking Analysis, and Evaluation Guidelines 1

Prepared for: Minnesota Department of Commerce, Division of Energy Resources 1

Prepared by: Illume Advising with subcontractors Indicia Consulting and Dr. Edward Vine 1

ACKNOWLEDGEMENTS 2

DISCLAIMER 2

Table of Contents 1

1.0 Executive Summary 5

1.1 Introduction 5

1.2 Key Findings 6

Behavioral CIP Program Definition and Classes 6

Program Benchmarking 6

1.3 Recommendations 13

2.0 Glossary of Key Terms 16

3.0 Study Introduction 18

Section 1: Literature Review and Benchmarking Study 21

4.0 Behavioral Programs Defined 22

4.1 Behavioral Program Definition 22

4.2 Behavioral Intervention Strategies 23

Benchmarking Analysis Findings 25

5.0 Calculus Family Findings 26

5.1 Feedback Programs 26

5.1.1 Asynchronous Feedback Programs 27

Residential Programs 31

Commercial Programs 31

5.1.2 Real-Time Pricing and Real-Time Data/Information Programs 32
Residential Programs 36
Commercial Programs 36
5.1.3 Benchmarking Programs 36
5.2 Diagnostics Programs 39
Residential Programs 39
Commercial Programs 41
6.0 Social Interactions Family Findings 43
6.1 Community-Based Programs 43
Residential Programs 47
Commercial Programs 48
6.2 Competitions 49
6.2.1 Game Competition Programs 50
Residential Programs 55
Commercial Programs 55
6.2.2 Non-Game Competitions (Commercial Only) 56
7.0 Cognition Family Findings 59
7.1 Education and Training Programs 59
7.1.1 Commercial Continuous Energy Improvement and Strategic Energy Management Programs 60
7.1.2 K-12 and Campus Programs 63
Section 2: Evaluation, Persistence, and Measure Life Overview 64
8.0 Recommended Evaluation Approaches 65
8.1 Establishing the Counterfactual 65
8.2 Considering Program Design in Selecting Evaluation Methods 68
8.3 Overview of Methods 70
Experimental Design 70

Quasi-Experimental Design 73

Preponderance of Evidence Approach 76

8.4 Methods Trade-offs 77

8.5 Cross Program Participation 77

Section 3: Energy Efficiency Behavioral Programs: Measure Life and Average Savings Methods Discussion 80

9.0 Behavior CIP Measure Life & Cost of Saved Energy 81

9.1 Context and Definitions: Integrated Resource Planning and Cost-Effectiveness Testing 81

Integrated Resource Planning and Achievement of CIP Goals 81

Utility Cost-Effectiveness Tests 82

Considering Behavioral Programs in Integrated Resource Planning 83

9.2 CIP Measure Life Consideration 83

9.3 The Standard Approach to Claiming Savings 85

9.4 The Average Savings Method Approach 88

9.5 The Avoided Decay Approach 90

9.6 Comparative Cost-of-Saved-Energy Under the Standard, ASM, and Avoided Decay Approaches 92

9.7 Accounting for CIP Efforts with Rolling Enrollment 93

9.8 Discussion and Recommendations 94

APPENDIX I: DETAILED METHODOLOGY 95

Approach to Defining Behavioral Programs 95

Approach to Defining the Taxonomy 95

Literature Review 102

Expert Interviews 102

Benchmarking Analysis 102

Reporting 104

References 105

Page intentionally left blank

1.0 Executive Summary

1.1 Introduction

The State of Minnesota Department of Commerce's Division of Energy Resources (DER) contracted with ILLUME Advising, LLC to complete a literature review study and benchmarking analysis of electric and gas utility-implemented Conservation Improvement Programs (CIP) that use behavioral techniques.[1] The overarching objective of this effort was to provide the State of Minnesota with the information necessary to make informed decisions regarding the design, evaluation, and claimed savings approaches for these programs.

To achieve this end, the literature review and benchmarking analysis had three primary goals:

(1) To develop a working definition of behavioral CIP efforts to be used in Minnesota;

(2) To identify, enumerate, catalogue, and benchmark success metrics associate with behavioral CIP efforts that have been evaluated to date (and are considered "evaluable"); and

(3) To determine practices associated with measurement, evaluation, and claiming[2] of behavioral CIP efforts to meet state and local resource acquisition goals.

To meet the first two goals, we first developed a working definition and taxonomy of behavioral CIP efforts. We then screened CIP programs that met our definition for evaluated savings. We included publicly available and evaluated programs that met our definition in our literature review, totaling nearly 170 studies including meta-analyses conducted in North America. After reviewing this research, we systematically categorized each program according to our taxonomy and then catalogued available success metrics associated with these programs.[3] Once the success metrics were collected, we compared the success metrics across behavioral programs. The team conducted 20 interviews with behavioral program thought leaders and program implementers were used to supplement our research and to identify publicly available data. Note, this work was designed to catalogue behavioral CIP efforts and no additional analysis or meta-analyses were conducted as part of this effort.

To meet the third goal, we consulted a number of sources to inform our findings and recommendations. These sources included: (1) publicly available literature on CIP evaluation best practices, (2) state filings and decisions related to CIP programs, (3) white papers pertaining to CIP measure life, and (4) industry experts knowledgeable about CIP filings in states who have extended CIP measure life.

We present our key findings and recommendations below.

---

[1] *For the purposes of this study, behavioral programs were defined as those energy efficiency programs that specifically use social science-driven behavioral strategies to influence energy consumption behaviors, and are evaluable for energy savings. Per the State of Minnesota, energy efficiency programs refers to measures or programs, including energy conservation measures, that specifically target consumer behavior, equipment, processes, or devices (2014 Minnesota Statues, 216B.241 Energy Conservation Improvement).*

[2] *Claimed savings includes both first year energy savings, and persistence of savings in subsequent years.*

[3] *See Appendix I for a list of success metrics collected.*

1.2 Key Findings

Behavioral CIP Program Definition and Classes

As a first step in our research, we developed a working definition of Behavioral CIP programs if the program deliberately applied behavior change models and approaches drawn from the social and behavioral sciences to affect energy use. Using the ACEEE Field Guide to Utility-Run Behavior Programs (2013) as our starting point, we developed a revised taxonomy that includes only those programs that have been successfully evaluated for energy savings.

After screening 170 studies for energy savings, we identified the following program families and associated program classes:

1. **The Cognition Family, including:** K-12 Schools; Continuous Improvement/Strategic Energy Management; and Benchmarking.

2. **The Calculus Family, including:** Diagnostics (On-Site and Remove); Asynchronous Feedback; and Real-time Feedback.

3. **The Social Interactions, including:** Community-Based and Competitions.

Program Benchmarking

Under our CIP definition and screening criteria, we identified a total of 58 CIP program models that have evaluated energy savings.

1. Most CIP efforts implemented and evaluated to date target the residential sector. We catalogued 35 evaluated residential programs and 23 evaluated commercial programs under our definition.

    a. We found few diagnostics programs (e.g., audits, assessments) that utilized the social science based methods outlined in our CIP definition.

    b. Members of some program classes do not attempt to claim direct savings, because they may be considered "educational" or used to drive participation in other programs (e.g., online energy assessments).

    c. Many programs that met the CIP definition (based on their tactics) did not provide evaluated savings in a manner that allowed comparisons with other programs (e.g., percent or per-premise savings).

    d. Some programs do not attempt to claim non-measure savings from the educational or behavioral component of the program. These include: Community-based programs, which typically only measure savings through existing retrofit programs; Diagnostic programs, which typically only claim savings through measures (though evaluations sometimes capture behavioral spillover).

CIP Programs generate a wide range of savings, varying dramatically by program class, target fuel, target sector, and target populations.

1) Among the residential sector, electric and gas savings range from 0%-6.5% and from 0.1%-2.2%,[4] respectively. Electric and gas savings range from -2%-22% and 0%-25%, respectively, for the commercial sector.

1.1 Among residential programs, the program classes with highest average **per premise** reductions include: Community-Based; Real-Time Feedback; and Competitions.[5]

1.2 Among residential programs, the program classes with the **greatest savings overall** are Asynchronous Feedback programs due to high participation levels.

1.3 Among commercial programs: the program classes with highest **average per premise reductions** include: Community-Based; Continuous Improvement; and K-12 schools.

1.4 Among commercial programs, few successful models have scaled to large population efforts or opt-out approaches due, in large part, to the need to implement targeted outreach to business owners and decision-makers.

Most of the behavioral program classes that met our criteria for having evaluated savings used some type of feedback, typically historical or comparative energy consumption.

- Providing feedback with social norms has become increasingly prevalent outside of Home Energy Reports programs: Competitions, opt-in feedback programs, and benchmarking programs often use these tactics.

- Fewer program classes used commitment, goal-setting and follow-through. Those that did include: Continuous improvement and strategic energy management (commercial), competition (games and non-games), and a select number of opt-in feedback programs.

- Few program classes routinely use non-price framing to encourage energy-saving actions, though we see a sizable opportunity in many behavioral program communications to frame opportunities in different ways (e.g., building energy ratings; comfort and convenience).

---

[4] *These ranges exclude savings achieved through community-based programs, because these programs only measure savings generated through existing retrofit programs. To avoid inflating the numbers in summary, we did not couple these values with other program efforts. Note some of the values in these ranges are net and gross, as many program classes did not report net savings.*

[5] *As we discuss further below, savings claimed for each program class are not always comparable to other program classes. Some do not measure net savings, and some measure savings through existing retrofit or rebate programs rather than conservation or non-measure-based savings.*

Evaluation and Policy

Few program classes have conducted rigorous experimental and quasi-experimental studies to estimate program savings. However, those program classes that have used these methods have strong evidence of savings.

The most rigorously evaluated CIP programs utilize experimental and quasi-experimental evaluation designs. While many use pre- and post-treatment analysis and preponderance of evidence approaches, we strongly caution readers in using these values for planning because the findings may not have effectively accounted for market conditions and other influences that would otherwise be controlled for in experiments and quasi-experiments.

- A select few classes have been evaluated using experimental or quasi-experimental methods to estimate net savings. These include: Asynchronous Feedback; and Real-Time Feedback. Notably, these program classes, however, have the greatest number of individual programs currently implemented by utilities. In particular, Asynchronous Feedback programs.

- The CIP efforts that have the greatest number of *third-party* (independent) evaluations include the following, in order of the number of studies completed: Asynchronous Feedback; Real-Time Feedback; Continuous Improvement; and Benchmarking.

- A number of CIP classes have reported savings associated with their efforts, however, most have *not undergone third party evaluations.* These include the following classes, in order of the number of studies completed: Games Competitions; Non-Game Competitions; Community-Based programs; and K-12. These program classes typically reported gross savings.

To date, only Residential Asynchronous Feedback programs have evaluated the persistence of savings both with and without continued treatment.

Few program classes have measured energy savings over multiple treatment years.

- Among those, only residential Asynchronous Feedback programs have evaluated persistent without continued treatment (i.e., after treatment stops).

- Some programs with shorter-term interventions rarely measure savings for an entire year. These include: Competitions (where the intervention may last for a few weeks or months); and Real-time Feedback (with a seasonal focus).

Most states implementing behavioral CIP efforts claim a one-year measure life for these programs. Residential Asynchronous Feedback programs are the only program effort that has successfully claimed measure life that exceeds one year, but this has not been widely adopted.

The only behavioral CIP class with a claimed measure life that exceeds one year is residential asynchronous feedback programs, however, the specific measure life used and the methods under which it is claimed vary by state and even within states.

  - o The ILLUME team was able to identify a total of five states where a multi-year measure life, including Connecticut, Colorado, New Hampshire, Nevada, and Washington.

- Most states that have extended the measure life of behavioral CIP efforts do so on an administrator-by-administrator basis. However, the Northwest is working toward a common measure life assumption and approach to claiming these savings in their Regional Technical Forum (RTF).

Table ES-1 summarizes our findings by sector and program class.

Table ES-2 outlined our findings on the evaluation approaches, and the levels of rigor associated with each program family and class.

Table ES-1. Summary of Program Savings Findings by Program Class

Family	Class	Most commonly used behavior-change tactics	Residential		Commercial		Comparability of results *among* behavioral programs (i.e., can you directly compare percent savings in this report across program classes?)
			Number included[a]	Savings Range	Number included[a]	Savings Range	
Cognition	K-12 Schools	In-person interactions; commitment	NA	NA	1	12%	Moderate (results are gross)
	Continuous Improvement/ Strategic Energy Management	Commitment; feedback; follow-through; in-person interactions	NA	NA	5	-2%-22% elec; 0%-25% gas	Moderate (results are rigorously obtained, but are generally gross, and may measure different results – e.g., all attributable savings vs. only O&M savings)
	Benchmarking	Feedback; framing	0	NA	2	1.1%-5% elec; 1.3%-7% gas	Moderate (difficult to isolate savings from benchmarking alone)
Calculus	Diagnostics (On-Site and Remote)	In-person interactions; framing	1	0%-6.5% elec	0	NA	Weak (results are based on savings from measure installations only; they do not assess conservation or non-measure changes)
	Asynchronous Feedback	Feedback; social norms	12	<0%-3.3% elec; 0.1%-2.2% gas	1	0.3% elec; 0.2% gas	Strong (net savings)
	Real-time Feedback	Feedback	6	0%-6% elec savings; 6%-30% peak demand	0	NA	Strong (net savings)
Social Interact-ions	Community-Based	In-person interactions; framing	6	12%-30% (all fuels)[b]	5	13%-31% (all fuels)[b]	Weak (results are based on savings from measure installations through existing programs; they do not assess (a) conservation or non-measure changes, or (b) savings among non-participating community members)

Family	Class	Most commonly used behavior-change tactics	Residential		Commercial		Comparability of results *among* behavioral programs (i.e., can you directly compare percent savings in this report across program classes?)
			Number included[a]	Savings Range	Number included[a]	Savings Range	
	Competition	In-person interactions; feedback; social norms; rewards	10	0.1%-14% elec (?)[c]	9	0-8% elec (?)[c]	Moderate (Savings are typically gross, and measured for a short duration)

a The number classified as behavioral and had evaluated percent savings. We discuss more programs than those counted in these columns, but many did not have evaluated savings that could be used in comparative benchmarking analysis. Those are mentioned in the text but not counted here.

b All community-based program percent savings in this report are gross savings achieved through existing retrofit programs

c Nearly all competition program savings are gross savings that were not verified by a third party.

Table ES-2. Summary of Evaluation Methods Findings by Program Class

Family	Class	Evaluation methods used	Rigor of Evaluations	Persistence evaluated?
Cognition	K-12 Schools	Pre/post billing analysis	Low, and typically gross savings	No
	Continuous Improvement/Strategic Energy Management	Pre/Post analysis billing analysis with engineering adjustments	Medium (rigorous engineering assessment, but difficult to utilize control group)	No
	Benchmarking	Pre/post billing analysis combined with net-to-gross ratio	Medium	Rarely
Calculus	Diagnostics (On-Site and Remote)	Engineering analysis (deemed savings or building energy software) combined with net-to-gross ratio (survey-based)	Medium (often net savings, but may rely on deemed savings)	No
	Asynchronous Feedback	Difference-in-differences billing analysis (utilizing Randomized Control Trial)	High (net savings and RCT)	Yes
	Real-time Feedback	Difference-in-differences billing analysis (utilizing matched comparison group)	High (net savings and matched comparison)	Sometimes
Social Interactions	Community-Based	Pre/post analysis or deemed savings/building energy software assumptions	Low, and typically gross savings	No
	Competition	Pre/post analysis	Low, and typically gross savings	Rarely

1.3 Recommendations

- **Consider the desired role of various behavioral program approaches and the ways they may be most effective in meeting the state's goals.** Each program family has demonstrated different success levels in achieving measurable savings. While this may be due, in part, to poorly designed evaluations, some program families may be better suited to general, portfolio support goals.
    - **The feedback family may be the most suited to energy efficiency resource standard (EERS) program models.** Feedback CIP efforts are the most rigorously evaluated family and have had the greatest success in demonstrating savings associated with their efforts. With an emphasis on billing and meter data information as the primary feedback mechanism, feedback programs provide a direct link between the treated consumer and their behavior (consumption). By design, they are well suited to experimental and quasi-experimental evaluations and thus lower risk programs.
        - For these programs, is important to note that savings per household are low, but the total savings may be high due to the number of households.
        - Other program classes have led to higher savings per household, but most have not had success achieving these savings at scale.
    - **The social interactions family, namely the competitions class, have the greatest likelihood of generating verifiable savings in the short-term, but have not been well evaluated.** Competitions may prove to be successful EERS programs, however, these programs need to go through more extensive research and development to (a) identify program designs capable of generating savings over longer periods of time, and (b) understand EERS program cross-participation. To date, most Competitions programs are implemented for weeks or months at a time.
    - **Some programs may be more appropriately used as channeling mechanisms that drive customers to other EERS programs.** Many online diagnostics programs and tools have not been evaluated to measure savings that can be directly attributed to their efforts. That said, EERS savings do not play a role in these programs – they are an outcome of the program participation. Savings above program participation have not been well evaluated.
        - It is important to note here that as research on household behavior becomes more sophisticated and utility customer data and information systems become capable of integrating usage and program participation data with predictive analytics, we expect these programs to become increasingly more effective at targeting and driving savings.
    - Some Cognition programs, despite having demonstrated high savings from behavioral, operations and maintenance changes, may be costly to implement due to their multi-pronged and high-touch intervention approach (e.g., training, feedback, executive meetings, follow-through). We recommend considering where these programs may be cost-effective compared with lower-touch, more scalable cognition and training offerings

- **Despite the growth of behavioral tactics in many EERS, we still see opportunity to integrate social science based insights into CIP and other efficiency programs.**
  - Only recently, have programs started to use social science based insights. The potential is great to make better use of key insights from the social sciences.
  - Few diagnostics programs utilize non-monetary framing: Most programs typically frame recommendations in terms of potential energy savings. There is an opportunity for multiple types of framing, including (a) improved choice architecture (fewer or more prioritized recommendations), (b) normative comparisons, such as energy ratings/benchmarking scores or "similar home" comparisons, (c) descriptions of benefits in terms beyond energy savings (i.e., multiple benefits), such as comfort, ease-of-use, maintenance, or home value.
  - Commitment and goal-setting appears to be an effective strategy in many program classes – Games/competition; Community-Based programs and Cognition programs. Program designers might look for opportunities to integrate this strategy into Calculus programs or other EERS programs.

- **Utilize CARD grant funding mechanisms to pilot and test program behavioral approaches that have the potential to support CIP goals.** Behavioral CIP efforts offer a significant opportunity to help regional utilities reach the state's goal of a 1.5% reduction in consumption. However, this class of programs still requires funding for research and development to further demonstrate their potential for generating sustainable savings. We recommend earmarking a portion of CARD grant funding opportunities to the development of these program approaches, including demonstrations and pilot testing of small-scale programs.

- **Evaluation requirements, and levels of rigor, should align with the intended role of the program.** While we strongly recommend that programs work to integrate experimental and quasi-experimental designs into their programs, we also recommend that this level of rigor only be required for those programs that aim to generate claimable program savings.
  - **For EERS efforts, quasi-experimental designs, at minimum, should be used for claimed savings.** For all behavioral efforts that are designed to generate savings as a unique resource program (vs. channeling to other programs), the State of Minnesota should require that such programs implement experimental designs (including quasi-experimental approaches) to estimate savings. In all cases, a comparison or control group should be drawn to establish the counterfactual.
  - **For channeling-focused programs, preponderance of evidence approaches may be sufficient.** For those programs that aim to generate effects throughout the portfolio, such as increases in program participation, methods that triangulate data sources may be sufficient to justify funding these types of programs.

- **Develop standardized reporting requirements for behavioral program success metrics to support comparative research and meta-analyses.** If the state is interested in continuing to study the relative success of behavioral CIP programs, we strongly recommend that the state require that behavioral programs provide a standard list of program metrics and information (building on regional and national efforts).
  - **Identify program performance information:** (1) net savings (using a comparison group), (2) adjusted net savings (net savings reduced by cross-program participation savings), (3) cross-program participation values, including the "lift" in participants and percent savings in other CIP programs, and (4) any discernable market effects or other impacts that may benefit the state as a result of the programs.

- **Express results on per-premise basis**: Report savings results on per-premise basis (e.g., average annual kWh savings) and as a percentage of baseline consumption (e.g. average percent savings).
- **Provide program design detail:** (1) the specific behavior change mechanism used, (2) the channels through which the customer is engaged, such as direct mail or mass media, (3) the intended outcomes of this engagement, specifically the targeted savings or participation goals, and (4) the supporting program theory or logic. Ideally, provide examples of customer-facing materials such as energy feedback, reports or online platform in evaluation documents.

2.0 Glossary of Key Terms

Automated Meter Infrastructure – technology which allows two way communication between a utility and an energy meter with an IP address; enables utilities to receive real-time data and in some cases control customer meters.

Billing Analysis – an analysis involving the use of regression models with historic billing data that calculates demand and energy savings.

Cohort – a group of program participants who began receiving treatment during the same period of time.

Commitment – a program feature which aims to influence participant behavior by asking participants to pledge either in a written or verbal format to adopt an energy saving actions.

Conservation Behaviors – actions taken energy users, which limit overall energy usage.

Critical Peak Pricing – an energy rate structure which includes an extra high rate for times when energy is most in demand.

Demand Response – Demand response programs encourage customers to shift the time of day when they use energy in order to reduce the system wide demand for energy.

Feedback – a program feature which aims to influence participant behavior by providing participants with information about how much energy they use over time.

Follow-through – a program feature which aims to influence participant behavior through subsequent interactions designed to remind the customer of the message delivered during the initial interaction.

Framing – a program feature which aims to influence participant behavior by presenting information to them in a way that may more effectively persuade them to take the desired actions.

Gross Savings – energy savings or changes in consumption resulting directly from program-promoted activities taken by participants regardless of extent or nature of the program influence on their actions.

Incremental savings – Savings which can be attributed to multiple energy efficiency programs.

In-person interaction – a program feature which aims to influence participant behavior by sending a program representative or partner to speak with a participant face-to-face.

Maintenance Practices – actions taken by equipment users over time to maintain optimal performance of energy-using equipment.

Matched Comparison – a quasi-experimental method which creates a one-to-one match between a population receiving treatment and a population not receiving treatment on the basis of similar qualities or characteristics.

Measure-Based Actions – the purchase of energy efficient equipment rebated by an energy efficiency program.

Measure Life – the number of years a measure or program is expected to generate energy savings.

Multi-pronged/stacked strategies – a program that employs two or more features designed to influence participant behavior.

Net Savings – energy savings or changes in consumption resulting from program promoted activities taken by participants, taking into account participant self-motivation or free ridership.

Net Adjusted Savings – net savings resulting from the program, excluding actions potentially rebated by a separate energy efficiency program.

Net Unadjusted Savings – net savings resulting from the program, including actions potentially rebated by a separate energy efficiency program.

Non-Measure Actions – either the purchase of energy efficient equipment not rebated by an energy efficiency program or behaviors adopted which limit overall energy use.

Operating Practices – the way in which an end-user regularly utilizes their energy-using equipment.

Opt in – a program design in which customers must enroll in a program in order to participate.

Opt out – a program design in which participants are automatically enrolled in a program, but may request to be unenrolled at any time.

Persistence –benefits over time resulting from an energy efficiency program or measure.

Randomized Control Trial – A program design which randomly assigns a subset of utility customers to a treatment group, which participates in the program, and a control group, which does not participate in the program. Savings are determined by measuring the difference in energy use between the two groups.

Randomized Encouragement Design - A program design which randomly assigns a subset of utility customers to a treatment group, which receives encouragement to participate in a program, and a control group, which does not receive encouragement to participate in a program. Savings are determined by measuring the difference in energy use between the two groups.

Rewards/Gifts – a program feature which aims to influence participant behavior by providing certain benefits in exchange for completion of desired actions.

Social norms – a program feature which aims to influence participant behavior by comparing them to members of their community.

Time of Use – an energy rate structure where energy rates differ throughout the day.

3.0 Study Introduction

The overarching goal of this study was to categorize, and benchmark, different types of energy efficiency behavioral programs against one another based on the success metrics of the program. Specifically, the following researchable areas were explored:

Objective 1: Systematically define, categorize, count, and enumerate types of behavioral programs used as energy efficiency resource programs.

1. Specific Aim 1: Develop a taxonomy of behavioral programs for inclusion in this study, including a definition of behavioral programs that outlines inclusionary and exclusionary criteria for behavioral programs in this study.
2. Specific Aim 2: Categorize and enumerate the behavioral programs that exist and/or have been implemented from 2010 through present within Minnesota and other states actively implementing behavioral programs.
3. Specific Aim 3: Catalogue the strategies that the selected behavioral programs utilize to drive behavior change.

Objective 2: Catalogue and compare success metrics across behavioral programs.

1. Specific Aim 1: Catalog and compare energy savings achieved by varying behavioral programs.
2. Specific Aim 2: Catalog and compare the persistence of savings achieved by different behavioral programs.
3. Specific Aim 3: Identify behavior-change tactics that may be instrumental in achieving greater savings levels.

Objective 3: Determine practices used by programs for measurement and evaluation of behavioral programs.

1. Specific Aim 1: Catalog how savings are measured, evaluated, and claimed by behavioral programs.
2. Specific Aim 2: Identify most common practices for evaluating behavioral programs.
3. Specific Aim 3: Determine which states have, if any, claimed a measure life that exceeds one year for behavioral programs and recommend an approach for the DER.

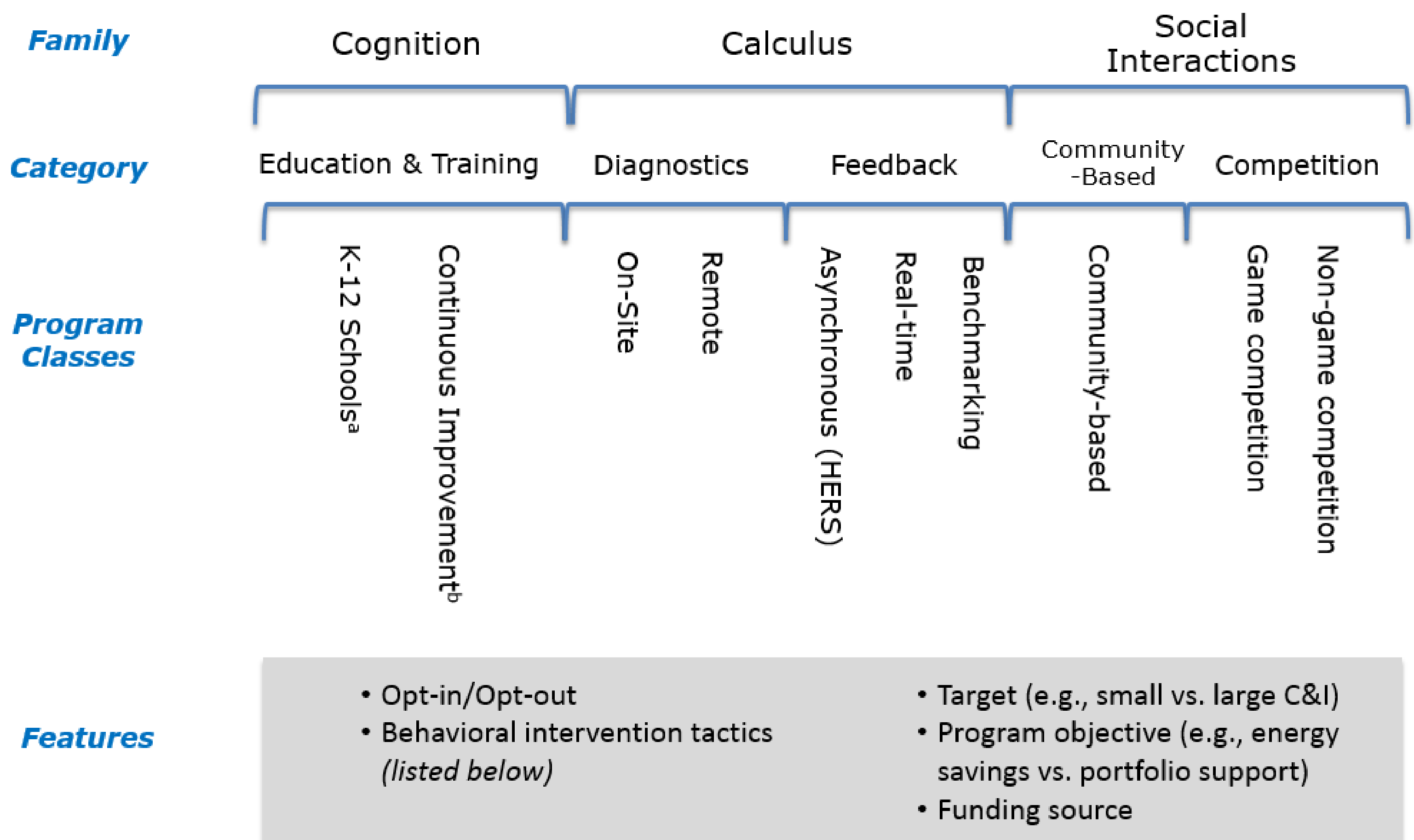
Programs were included in this study if: (1) they met the definition of a behavioral CIP program, namely if the program deliberately applied models and approaches drawn from the social and behavioral sciences to affect energy use; and (2) the program was evaluable for energy savings in a manner that could be compared across programs. Our approach differs from the ACEEE definition of behavioral programs, in that we specifically sought to include only those programs for which evaluative data related to savings existed. There are many programs that use behavioral tactics, but that have not been rigorously evaluated. These were excluded from this benchmarking analysis, given objective 2, above (to catalogue and compare metrics across programs).

In addition to restricting our study to include programs that met the above definition, we also developed a program taxonomy to define different program classes (described in more detail in the Methodology appendix). This taxonomy served as a guide for this study, but not as strict exclusion or inclusion criteria for specific programs. As we conducted our literature review, we learned that there is a dearth of savings-related data for certain program classes. As such, in the Findings section of this report, some program classes have more robust findings, while other program classes have limited findings.

In summary, the following steps were completed to accomplish the literature review and benchmarking analysis. A detailed overview of the methodology is provided in Appendix II.

- Define Behavioral Programs: Programs met our definition of "behavioral" if they specifically used behavioral strategies to influence behavior change related to energy consumption, and if they were evaluable for energy savings.
- Define Behavioral Program Taxonomy: We created taxonomies of behavioral programs for both the residential and commercial sectors. Originally, our taxonomy was very similar to that proposed by ACEEE. However, as we conducted our literature review, we found the below (Figure 1), to be more appropriate. For example, an original classification in our taxonomy included both community-based and peer-to-peer programs. However, because peer-to-peer engagement strategies are common in community-based programs, we do not report on peer-to-peer as stand-alone programs.
- Literature Review: During January and February, 2015, we conducted a literature review of behavioral programs that had been implemented during 2010 or thereafter and also asked experts in the field for their suggestions. This review included research across multiple sources, including CALMAC, published papers, peer-reviewed white papers, third-party evaluations, implementer-led studies, and conference proceedings, including those available from ACEEE and the International Energy Program Evaluation Conference (IEPEC). As we identified new programs, we first ensured that they met our inclusion criteria. Programs that did meet the study inclusion criteria were then classified per the program taxonomy. In addition to identifying individual behavioral programs, we also reviewed meta-analyses of behavioral programs. In total, we reviewed approximately 170 evaluations, studies, and meta-analyses.
- Expert Interviews: We conducted 20 interviews with program administrators and evaluators of behavioral feedback programs. We initially contacted those individuals via email to inquire about particular programs (or program classes) and whether evaluations of those programs had been completed. We then conducted telephone interviews to follow-up on the study recommendations. In addition to these interviews, we interviewed five Minnesota-based program administrators to discuss behavioral CIP efforts in the state.
- Benchmarking Analysis: We developed several metrics that we used to benchmark behavioral programs against one another. It is important to note that the documentation of behavioral programs often only captured a selection of these metrics. As such, metrics that were publicly available were included. The primary metric that we used to benchmark programs was evaluated savings (expressed as a percentage of baseline consumption).

Figure 1. Program Taxonomy for the Residential and Commercial Sectors.



[a] *We consider K-12 Schools programs to be a "Commercial" offering based on where the program is delivered, though savings may be achieved in the school or student homes*

[b] *Continuous Improvement (also known as Strategic Energy Management) is a commercial-only offering*

Section 1: Literature Review and Benchmarking Study

The first section of this report contains findings from the behavioral program benchmarking analysis, including our definition and screening criteria, and detailed findings by program class. Section Two of this report contains evaluation findings, evaluation best practices, and a discussion of policy issues related to claiming CIP program savings. Readers may wish to cross-reference our overview of evaluation design in Section 2 while reading Section 1, as the ability to compare benchmarking findings across program classes depends on comparability of evaluation approaches (e.g., the use of a comparison or control group to estimate net savings).

4.0 Behavioral Programs Defined

In this section, we elaborate upon the behavioral program definition that we developed for use in this study. Programs were included in this study if they met this definition. We also describe specific behavioral strategies that behavioral programs may use, and that we specifically searched for in reviewing behavioral programs.

4.1 Behavioral Program Definition

In 2012, SEE Action defined behavioral programs as "those that utilize strategies intended to effect consumer energy use behaviors in order to achieve energy and/or peak demand savings. Programs typically include outreach, education, competition, rewards, benchmarking and/or feedback elements" (Todd, Stuart, Schiller, & Goldman, 2012). Recently, the energy efficiency community has coalesced around a tighter definition of behavior programs, as defined during the CA IOU Behavior Summit in the summer of 2013 and the AAAS/NYSERDA Behavior Summit in the summer of 2014. Because of this, we recommend that there be made a distinction between earlier 'vernacular' behavior programs, which were often the product of trial and error, and newer 'modern' behavior programs that follow the recommendations coming out of publications and convenings. Unfortunately, the timescale required for a program design to be developed, approved, implemented, and evaluated means that few, if any, programs that both fulfill 'modern' requirements and have energy savings data that have made it through the pipeline.

In order to determine behavioral program inclusion in this study, we use the 'modern' definition of behavior. Specifically, we adopted the following two criteria to determine behavioral program inclusion:

1. Adopt the **California working definition** of "behavioral programs" developed in 2014 at a workshop with the IOUs, CPUC and other interested stakeholders. At this workshop, the IOUs recognized that: "All existing DSM and energy efficiency programs involve behavior, and [should apply] [benefit from] insights from social and behavioral sciences for greater impact and deeper savings." However, for implementation purposes, the CPUC defined a separate funding category of new behavior-based innovations as programs: "that deliberately apply models and approaches drawn from the social and behavioral sciences to affect energy use." To differentiate these programs from traditional, incentive-based demand-side management (DSM) programs, the IOUs et al. further classified these new behavior-based innovations as efforts or interventions that use social science theory, and identify energy usage behaviors that are intended to be changed (see appendix I for a full definition) (Ignelzi,, Randazzo, Dethman and Lutzenhiser, 2013).
2. The program class or category must be **evaluable**. The programs we set out to benchmark must be evaluable in the sense that energy savings impacts must be:
   (a) quantified through accepted industry methods;
   (b) quantified at the level specified in our taxonomy; and
   (c) quantified in a manner that allows comparisons across programs (i.e., average percent saving).[6]

---

[6] *The latter part of this evaluability criteria is another reason for benchmarking by program rather than by behavioral tactic – in some cases, a program may incorporate multiple tactics (behavioral and non-behavioral), and it may not be possible to isolate results by tactic.*

Note that this criteria does not preclude programs whose primary objective is not generating energy savings, as long as energy savings are estimated in some manner. Programs whose primary objective is market transformation (e.g., increasing market share of energy-efficient appliances) or portfolio support (e.g., promoting other energy-efficiency rebates offered by a utility) can be included in benchmarking insofar as energy savings impacts are evaluated. Still, some exclusions resulted from this rule, including (a) awareness-generating campaigns which measure "outputs" like number of attendees or flyers distributed rather than impacts; and (b) online forums, for which no evaluation model has been developed.

To further differentiate behavior-based innovations from traditional, incentive-based demand-side management (DSM) programs, the IOUs et al. further classified these new behavior-based innovations as efforts or interventions that:

1. Identify energy usage behaviors that are intended to be changed;
2. Identify which social science theory or combination of theories the intervention is drawing upon;
3. Deploy behavior intervention strategies *(listed in next section)*;
4. Utilize messaging strategies grounded in behavioral and cognitive sciences;
5. May be evaluated using experimental design, quasi-experimental design or other evaluation methods approved by CPUC; and where
6. Outcomes are typically measured on an ex-post basis, using approved evaluation methods; however, in some cases, forecasted metrics may be used.

4.2 Behavioral Intervention Strategies

A behavioral program may use a wide range of strategies to effect change, including social science-based strategies, as outlined below, as well as more "traditional" strategies, such as pricing and incentives or information-based outreach. As such, we also specified the types of behavioral intervention strategies that were prioritized for inclusion in this study. These strategies served as prerequisites for a program's inclusion in our literature review and benchmarking effort. This list combines tactics enumerated and defined in the California whitepaper, "Paving the Way for a Richer Mix of Behavioral Programs,"[7] and the "ACEEE Field Guide to Utility-Run Behavior Programs." These strategies include the following:

- Commitment (including goal-setting)
- Feedback
- Follow-Through
- Framing (this includes "nudges" like strategically setting default options)
- In-person interactions (through a trusted community messenger or social diffusion)
- Rewards or gifts
- Social norms
- Multi-pronged or "stacked" strategies (using two or more tactics)

Some of these behavior change intervention strategies may be present in traditional, incentive-based DSM programs. However, per the California working group definition of behavioral programs

[7] *Ignelzi, Patrice; Jane Peters; Katherine Randazzo; Anne Dougherty; Linda Dethman; Loren Lutzenhiser. "Paving the Way for a Richer Mix of Residential Behavior Programs." Prepared for the California Investor-Owned Utilities: Pacific Gas & Electric, Southern California Edison, San Diego Gas & Electric, and Southern California Gas. San Francisco, CA. June 2013.*

(provided above), we only included those programs that rely predominantly on these intervention strategies rather than "traditional" strategies (i.e., simply providing information and money).

Two common features of energy efficiency programs that are not part of this core list of underused, social science-based strategies include:

- Information and outreach (e.g., direct marketing telling consumers that installing a specific appliance or weatherizing could help them save $X on their electric bill)
- Price-based incentives (e.g., rebates or time-of-use rates)

For the purpose of this study, we included programs with these elements, insofar as they also rely (to a significant extent) on other social science-based intervention tactics from the list above. For example, if a community-based program relies mainly on enhanced incentives to increase uptake in another program (i.e., rebates, audits, or home energy upgrades), and outreach is primarily focused on that increased incentive, we excluded it from our study. However, some community-based programs use energy information feedback, peer comparisons and social interaction, thus qualifying for inclusion in our study. Similarly, for years, utilities have run commercial programs aimed at training building operators, changing operating and maintenance practices, and providing detailed feedback on energy information (usually meter data) aimed at identifying opportunities to modify operating practices. These programs are typically not described as "behavioral", though they use many of the strategies described above, and as such, would meet the inclusion criteria for this study.

Benchmarking Analysis Findings

In the following sections, we present key study findings for the benchmarking analysis. The section consists of several subsections, which align with the program categories identified in the taxonomy. As noted earlier in this report, for some program classes, we identified many programs that met our inclusion criteria (i.e., asynchronous feedback). For other program classes, few programs met our inclusion criteria, and thus, findings related to those program classes are limited.

Within each program category, we begin by presenting results related to the residential sector. These are then followed by results for the commercial sector within the same program category. Benchmarking results are presented for all program classes that fall within that category, in order to facilitate comparisons among program classes within the same category. In addition to presenting our benchmarking results, we also highlight notable program examples in text.

For the savings metrics presented in this section, where possible, we present “net unadjusted savings”. Nearly all programs are required to evaluate and report net savings for the purpose of claiming savings. However, not all programs are required to remove the portion of net program savings that may also be claimed by other utility rebate programs (sometimes referred to as, “joint savings”, “incremental savings” or “double-counted programs”). To enable comparisons across programs and program classes, we have chosen to present “net unadjusted savings” where possible, instead of “gross savings” or “net adjusted savings.”

Net unadjusted savings are defined as the whole of net program savings that can be attributed to program efforts, per the evaluation methods. These savings may include savings that participants achieved through the installation of measures rebated through other DSM measures. Typically, the other DSM programs claim 100% of these savings. However, because the control or comparison group is also installing rebated measures at some natural rate of participation, not all of the savings achieved through the installation of program measures need to be removed from net program savings to avoid double-counting. Only the portion of savings that is “incremental” – over and above the level achieved by the control or comparison group – will be double-counted. In practice, these incremental savings are fairly small among the programs that measure them, representing approximately 6% of net electric program savings, and 4% of net gas savings among Home Energy Report programs[8], which are the only behavioral programs to routinely estimate these savings. Moreover, since the decision of what program(s) can claim credit for these “joint” or “incremental” savings is a current policy topic, we elected to report net unadjusted savings in this study to provide readers with an understanding of net savings these programs are capable of achieving.

[8] *This is a weighted average across all HER programs and cohorts in this study – weighted by the number of treatment customers in the cohort.*

5.0 Calculus Family Findings

We present our findings for the Calculus family of programs in this section, defined as programs that provide energy-related information that customers need to make economically rational decisions about energy use (Mazur-Stommen & Farley, 2013). Of all behavior program families, Calculus programs are the most widely implemented and the most thoroughly evaluated. This family includes the following program categories, listed in order of the prevalence of evaluated programs: (1) Feedback (residential and commercial sectors); and (2) Diagnostics (residential and commercial sectors).

5.1 Feedback Programs

ACEEE defines behavioral feedback programs as those that "provide customers with information to encourage shifting of loads to off-peak periods and/or to encourage lower levels of overall consumption" (Ehrhardt-Martinez, Donnelly, & Laitner, 2010). Within the Feedback program category, our benchmarking analysis compares the following program classes, in order of evaluated data: (1) Asynchronous Feedback; (2) Real-time Pricing and Real-time Data/Information; and (3) Benchmarking (commercial only).

Feedback programs operate under the assumption that customers who are well-informed, or made aware of, their energy use will be motivated to reduce their consumption. For customers to adopt new behaviors related to their energy consumption, they must: (1) understand the information being presented to them, (2) believe that they are capable of taking actions, and (3) through the use of behavioral strategies, be "nudged" into taking specific actions to reduce their energy consumption (such as turning off lights when not in a room, unplugging electronic devices, or installing energy-efficient appliances) (Alcott & Mullainathan, 2010).

Within this program category, we found the following:

- Twenty-nine residential and five commercial feedback programs implemented between 2008-2012 met our criteria of having both behavioral interventions and evaluated savings.
    - Asynchronous Feedback, particularly residential Home Energy Reports (HERs), are the most widely implemented and robustly evaluated programs in this family. These programs generate roughly between 0.9% and 2.2% net electric and between 0.3% and 1.6% net gas savings.[9]
    - Real-time feedback programs were the second most commonly evaluated, typically providing daily or hourly feedback. Among residential real-time feedback programs, net unadjusted electric savings ranged from 0% to 3.1%.
    - We did not identify any real-time pricing and real-time data/information programs that met our criteria for use of CIP intervention tactics for the commercial sector. Within the commercial sector, real-time feedback is typically not delivered as the primary component of a program, as it is for the residential sector. Instead, real-time feedback is often part of commercial diagnostics programs, where it is bundled with additional information and recommendations, or as part of Continuous Improvement programs, where it is bundled with in-person interactions, training, commitment, goal setting, and follow-through.

[9] *Net unadjusted savings refer to changes in energy use attributable to a particular energy efficiency program.*

- Of all Feedback programs reviewed, the most common social science interventions are used, listed in order of their prevalence in programs: (1) feedback; (2) social norms; and (3) rewards.
- The most common evaluation methods used for this program class include randomized control trials with opt-out design for asynchronous programs, and matched comparison design for opt-in real-time feedback programs.

5.1.1 Asynchronous Feedback Programs

Asynchronous feedback programs are defined as programs that provide feedback to customers after energy consumption has occurred (Ehrhardt-Martinez et al., 2010). Of the evaluations we reviewed, the most common program model included feedback delivered in the form of Home Energy Reports (HERs) for residential customers. This feedback is most frequently delivered in separate mailings, but may also be included in monthly energy bills. Some programs also provided feedback via email or online portal. Only one residential feedback program coupled feedback with reward strategies. There are very few asynchronous feedback programs for the commercial sector.

Residential asynchronous feedback programs' generate net unadjusted savings ranging from 0.9-2.2%, though savings are dependent on a number of factors. Baseline energy usage is the factor that most impacts asynchronous feedback programs' ability to generate savings. Higher baseline consumption households generally achieve greater savings on an absolute and percentage basis as compared to lower consumption households (Fuller et al., 2010; Ashby et al., 2012). Greater savings have also been reported among households with fewer occupants, smaller square footage, and older heads of household (Davis, 2011). Asynchronous feedback programs also demonstrate greater relative and absolute savings when targeting electric customers. It should be noted that we describe issues of persistence in a later section of this report.

Researchers have shown that residential asynchronous feedback programs can be enhanced with the addition of behavioral interventions, such as goal-setting, commitments, competitions, and games (Ehrhardt-Martinez et al., 2010). As an example, in 2013, National Grid Rhode Island piloted HERs plus rewards. This resulted in an additional 0.98 percentage points in savings for electric and 0.44 additional percentage points in savings for gas, in addition to those savings gained by the HERS without rewards (Illume Advising, 2014).

Due to the sheer number of residential asynchronous programs in implementation, we have not benchmarked all efforts. Instead, we focused on 12 notable programs, across several states, implemented between 2008-2012. Programs include those delivered in Minnesota (Xcel Energy), California (Sacramento Municipal Utility District, Pacific Gas & Electric), Illinois (Ameren, Commonwealth Edison), Washington (Puget Sound Energy), Massachusetts (National Grid, and Eversource (formerly NSTAR and Western Massachusetts Electric Company), and Rhode Island (National Grid). As noted earlier, there are very few commercial programs implemented at present and fewer have been evaluated.

The tables below summarize all key benchmarked metrics for asynchronous programs included in this study. Following the tables, we describe residential and commercial asynchronous programs.

Table 1. First-Year Net Unadjusted Savings Associated with Asynchronous Feedback Programs for Residential Sector

Program	Behavioral Strategies	Design	Participant n	# of Cohorts with First-Year Savings [b]	Average Duration for Savings Estimation (years)	Net Unadjusted Electric Savings	Net Unadjusted Gas Savings
Ameren IL Behavioral Modification	Feedback, social norms	Opt-out	198,183	3	0.67	0.9%-1.3%	0.4%-1.0%
ComEd HER (IL)	Feedback, social norms	Opt-out	259,261	3	0.75	1.2%-1.7%	NA
CUB Energy Saver (IL)	Feedback, social norms, rewards	Opt-out	8,793	1	1.00	2.0%	NA
MN Enerlyte	Enhanced bill cell phone application	Opt-out	24,326	1	1.00	2.2%	NA
NGRID RI Statewide	Feedback, social norms, rewards	Opt-out	269,174	6	0.50	-2.2%-1.6%	0.3%-0.5%
PG&E HER (CA)	Feedback, social norms	Opt-out	542,411	6	1.08	0.9%-1.5%	0.4%-0.9%
Puget Sound Energy HER (WA)	Feedback, social norms	Opt-out	31,618	1	1.00	1.7%	1.2%
SMUD HER (CA)	Feedback, social norms	Opt-out	100,347	3	1.00	1.6%-1.8%	NA
Xcel HER (MN)	Feedback, social norms	Opt-out	32,762	1	1.00	2.1%	0.6%
NGRID HER (MA)	Feedback, social norms	Opt-out	653,908	12	1.00	1.0%-1.7%	0.5%-1.2%
NSTAR HER (MA)	Feedback, social norms	Opt-out	144,739	5	0.73	1.5%-1.6%	1.0%-1.6%
WMECo Western Mass Saves	Feedback via paper HER and online portal	Opt-out	92,485	3	0.61	0.0%-1.9%	NA

[a] Note: We did not report opt-out rates for these programs, because they are typically quite small.

[b] Note: Number of cohorts with first-year savings is only relevant for Home Energy Reports. Other program classes either do not have waves or cohorts, or report them separately (whereas here, we collapse the number of cohorts with first-year energy savings).

Table 2. First-Year Net Unadjusted Savings Associated with Asynchronous Feedback Programs for Commercial Sector

Program	Behavioral Strategies	Design	Number of Treated Customers	Average Duration for Savings Estimation	Net Unadjusted Electric Savings	Net Unadjusted Gas Savings
PG&E BER	Feedback; social norms	Opt-out	15,328	10 mos.	0.3%	0.1%

Below, we graphically represent electric and gas savings versus cohort maturity for residential HER programs. Generally, as cohort maturity increases, the percent savings associated with these programs also increases.

Figure 2. Scatterplot of Electric HER program savings versus cohort maturity (among cohort with 2 or more years of continued program treatment)

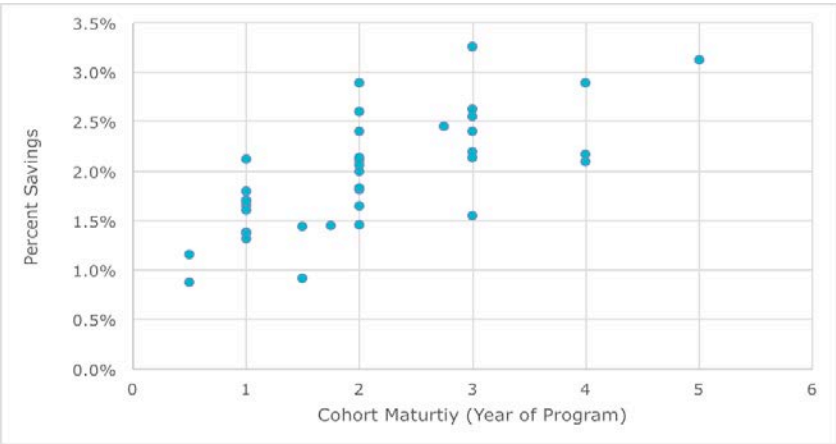
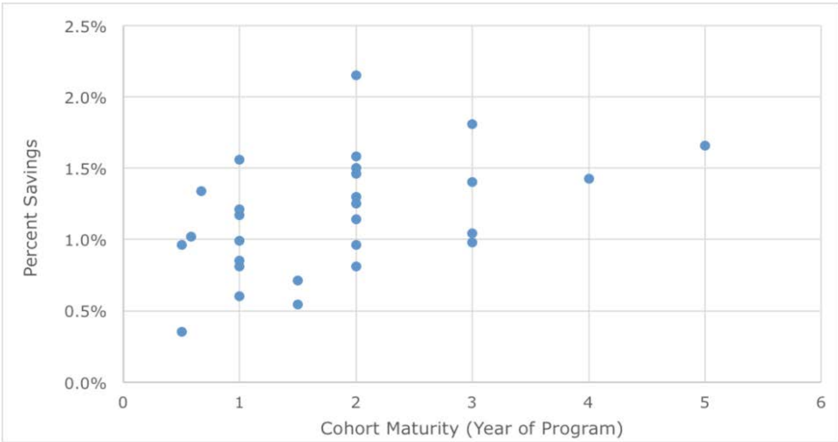


Figure 3. Scatterplot of Gas HER program savings versus cohort maturity (among cohort with 2 or more years of continued program treatment)



Residential Programs

The most common asynchronous feedback program models implemented are Home Energy Reports (HERs). Of HER programs, the most widely implemented is the Opower HER program (Alberta Energy Efficiency Alliance, 2014),[10] though other implementers include Tendril, C3 Energy, and Direct Options. HERs are different from other asynchronous online feedback platforms, in that they empower customers to make energy decisions by providing energy usage information and billing information, layered with interpretative and explanatory material that is actionable. In contrast, older asynchronous feedback platforms simply provided basic energy usage information about homes, without interpretation or specific action items that customers could take to decrease energy use. HERs vary in the type of information that is presented to the customer, the frequency with which it is provided (such as monthly, bi-monthly, quarterly, or seasonally), and how that information is provided (for example, on paper, via an online web portal, or through a mobile application). Several studies have found that more frequent feedback (i.e., monthly or bimonthly) results in higher energy savings, as compared to feedback that is delivered on a quarterly basis (Alcott, 2011; Abrahamse, Steg, Vlek, & Rothentagger, 2005; Alberta Energy Efficiency Alliance, 2014). For example, a study of HERS delivered by the Sacramento Municipal Utility District found average annual electric savings of 2.4% among customers who received monthly HERS, as opposed to average annual electric savings of 1.4% among customers who received quarterly HERs (Integral Analytics, 2012). Similarly, an evaluation of Puget Sound's Home Energy Report program found higher electric savings among current recipients of monthly HERS as compared to quarterly HERS; however, these findings were not statistically significant. Gas savings among current recipients of HERS were significantly higher among those who received monthly, as opposed to quarterly, feedback (DNV-GL, 2014a). It should be noted that savings tend to be highest immediately following delivery of feedback. As described in a 2012 CADMUS report, there are three savings phases: the first occurs during the six to 12 months of program participation, and savings increase rapidly during this phase. During the second phase, from the next 12 to 24 months, savings continue to increase, but at a lower rate as compared to phase one. During the final phase, typically in program years 3 and 4, savings begin to level off (Khawaja and Stewart, 2014/2015). Opower distinguishes its brand of HERs from older types by their use of social norms as an intervention tactic, which is implemented by comparing the participating customers' energy consumption to similar households. HER programs may also include other components, such as benchmarking a household's energy use over time, providing energy savings tips, and providing estimates of how much money could be saved by taking certain actions (such as installing energy-efficient lighting, or replacing an older appliance with a more efficient appliance).[11] As noted earlier, feedback programs may achieve greater savings when coupled with other behavioral strategies. As an example, the CUB Energy Saver program included rewards and reported net unadjusted savings of 2.0%.

Commercial Programs

Asynchronous feedback programs may also be implemented in the commercial sector. However, we only found one program that met our inclusion criteria, and discuss that briefly, below.

---

[10] *While OPower has the greatest reach, currently working with 80 utilities in 15 countries, other companies, such as Tendril, also provide Home Energy Report platforms (Alberta Energy Efficiency Alliance, 2014).*

[11] *Evaluations of Home Energy Report programs do not disentangle savings effects of these individual Home Energy Report components. As such, all savings estimates presented in this section are for Home Energy Report programs in their entirety (and not for individual Home Energy Report components).*

Example 1: Pacific Gas and Electric's (PG&E) Business Energy Report (BER) program for small- and medium-sized businesses.[12] Similar to Home Energy Reports, the BER provides a bimonthly feedback report to businesses on their energy usage, includes social norming behavioral strategies, and also provides customers with recommended energy-saving tips. The energy usage reports depict energy usage based on time, cost, and weather. A billing analysis conducted in 2014 revealed that net unadjusted electric savings were 0.3% (as compared to baseline), while net unadjusted gas savings were 0.1%.

5.1.2 Real-Time Pricing and Real-Time Data/Information Programs

Real-time pricing and real-time energy data programs are defined as those programs that provide real-time feedback on energy pricing and energy consumption, respectively. We identified several real-time feedback programs that met our criteria for use of behavioral intervention tactics within the residential sector. However, we did not find any programs within the commercial sector. Among real-time pricing programs, feedback was most frequently provided via an online portal or in-home display. For real-time data/information programs, hourly or daily energy consumption feedback was most frequently provided via an online portal. Three of the real-time feedback programs that we reviewed also included commitment and goal setting behavioral strategies.

Residential feedback programs generate net unadjusted electric savings ranging from 0%-3.1%%. We did not identify real-time feedback programs within the commercial sector that met our criteria, and thus, do not report savings for this sector. Two different meta-analyses have found strong effects of price signals on the timing of electricity consumption (Faruqui & Sergici, 2010; Newsham & Bowker, 2010). To be most effective, real-time pricing and data programs should integrate other behavioral strategies (Ehrhardt-Martinez et al., 2010). Similarly, Buchanan, Russo, & Anderson (2015) concluded that successful in-home device (IHD) rollout requires the "human factor" to keep customers engaged. Buchanan et al. define this as "the components and processes involved in consumer engagement with in-home devices." While Buchanan et al. (2015) do not present their own evaluation results, they do state that "IHDs cannot reduce energy consumption by themselves, but that this depends on human interaction and action." The authors argue that: consumers must be engaged and become interested in energy reduction; motivation should be primed through the delivery of rewards, as opposed to a focus on cost; energy information must be provided that is easily comprehensible; and that a "one size fits all approach for IHDs cannot be justified."

We report on residential real-time feedback programs with implementation dates as early as 2011. Programs include those delivered in Minnesota (Minnesota Power), California (Sacramento Municipal Utility District (SMUD), Southern California Edison), Maryland (Baltimore Gas and Electric), and Arizona (Tucson Electric Power). As noted earlier, we did not identify real-time feedback programs that met our criteria for the commercial sector.

Table 3 summarizes all key benchmarked metrics for real-time feedback programs included in this study. Within this table, we have included summer peak load reduction savings. This metric is only relevant for this program class, and as such, is not reported on for other program classes. Following Table 3, we describe real-time feedback programs.

---

[12] *PG&E generally defines small businesses as those that consume less than 20kW; medium businesses are those that consume between 20kW and 200kW.*

Table 3. Participation and Savings in Residential Real-Time Feedback Programs with Real-Time Pricing

Program	Behavioral Strategies	Design	Participation Rate[a]	Participant (n)	Average Duration for Savings Estimation	Net Unadjusted Electric Savings	Summer Peak Load Reduction (%)[b]
SMUD SmartPricing Options: CPP	Feedback via IHD, web portal	Opt-in	18.2%	1,651	12 CPP events	NA	26%
SMUD SmartPricing Options: CPP	Feedback via IHD, web portal	Opt-out	95.9%	701	12 CPP events	NA	12%
SMUD SmartPricing Options: TOU	Feedback via IHD, web portal	Opt-in	17.5%	2,199	4 mos.	NA	13%
SMUD SmartPricing Options: TOU	Feedback via IHD, web portal	Opt-out	97.6%	2,018	4 mos.	NA	6%
SMUD SmartPricing Options: CPP	Feedback via Web portal only	Opt-in	18.8%	223	12 CPP events	NA	22%
SMUD SmartPricing Options: TOU	Feedback via Web portal only	Opt-in	16.4%	1,229	4 mos.	NA	10%
SMUD SmartPricing Options: TOU+CPP	Feedback via IHD, web portal	Opt-out	92.9%	588	4 mos.	NA	8% summer peak; 13% critical peak
Edison SmartConnect: IHDs	Feedback via IHD with real-time cost	Opt-in	NR	183	6 mos.	6% in first 60 days; 0 thereafter	NA

Table 4. Participation and Savings in Residential Real-Time Feedback Programs without Real-Time Pricing

Program	Behavioral Strategies	Design	Participation Rate[a]	Participant (n)	Average Duration for Savings Estimation	Net Unadjusted Electric Savings	Summer Peak Load Reduction (%)[b]
Edison SmartConnect: Budget Assistant	Goal-setting and notifications tools in “My Account”	Opt-in	NR	117,377	6 mos.	0.92% (but diminish over time)	NA
Edison SmartConnect: IHDs	Feedback via IHD with*out* real-time cost	Opt-in	NR	163	6 mos.	3% in first 30 days; 0 thereafter	NA
Minnesota Power AMI Pilot	Feedback via online portal with either daily or hourly consumption	Opt-in	4.7%	2,523	1 year	0%	NA
Minnesota MyMeter	Online portal/app with feedback, comparative usage, goal-setting, notifications	Opt-in	9%-16%	14,156	1-3 years	1.8% - 2.8%[c]	NA
National Grid: EmPower (RI)	Feedback via online portal; communicating outlet (72%) or thermostat (26%)	Opt-in	<1%[d]	90	1 year	1.7%	Range from 30% savings to 19% increase in consumption
Tuscon Electric Power: Power Partners	Online portal with AMR data, recommendations, goal-setting, challenges, forum	Opt-in	5-6%	1,521	8 mos.	1.2%-3.1%[e]	NA

[a] Opt-in rate for opt-in programs, and (1-opt-out rate) for opt-out programs

[b] For the cohorts with critical peak pricing, the percent reduction is during critical peak periods.

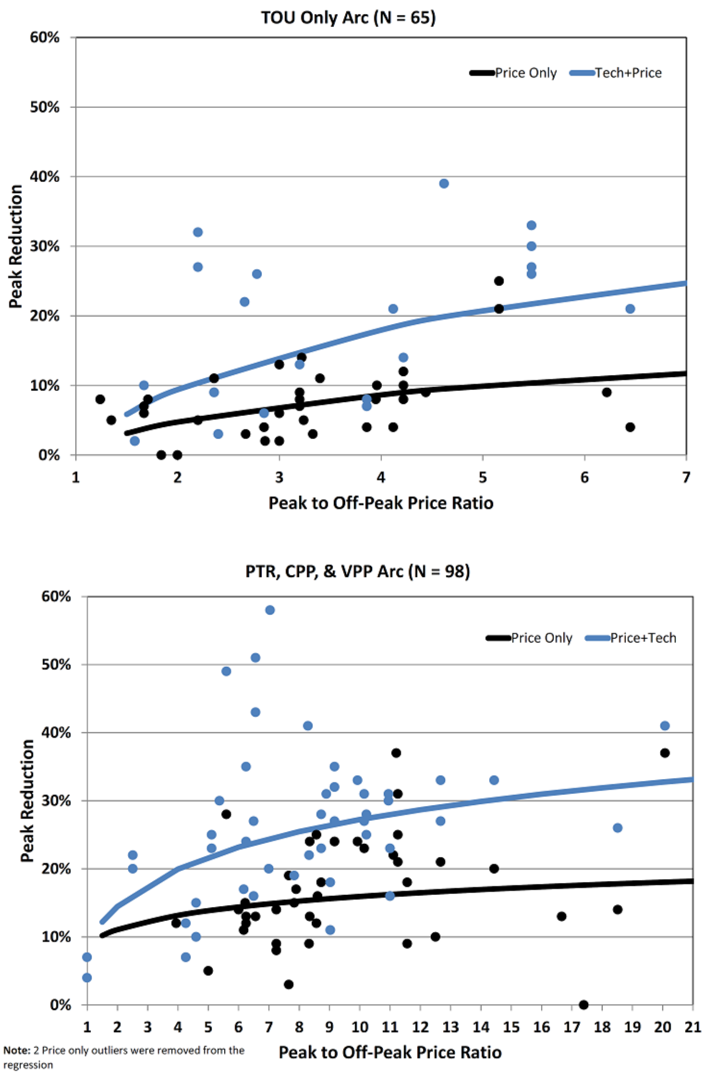
[c] Results cover a three-year period, so they may not be comparable to first-year savings results for other programs

[d] It is unclear how program recruitment was targeted or marketed the EmPower pilot was. If recruitment was targeted to specific customers rather than marketed to all, the true participation rate may be higher.

[e] Billing analysis covers an eight-month period.

We did not attempt to catalog real-time pricing programs where pricing is the primary intervention strategy. As noted above, two different meta-analyses have found strong effects of price signals on the timing of electricity consumption (Faruqui & Sergici, 2010; Newsham & Bowker, 2010). Ahmad Faruqui and colleagues at the Brattle Group have compiled peak savings results from multiple dynamic pricing trials, and plotted the relationship between the peak-to-off-peak pricing ratio and savings for programs that offered some type of enabling technology, versus those that do not. In general, they found that enabling technology (e.g., in-home displays, energy orbs, smart thermostats[13]) increases the price response for a given peak-to-off-peak pricing ratio (Faruqui, 2013).

Figure 4. The Arc of Price Responsiveness for Time-of-Use Pricing and Dynamic Pricing



---

[13] *In some cases the enabling technology served to automate the price response rather than provide feedback, while some enabling technology is purely feedback. The authors did not examine the difference in savings between dynamic pricing enabling by automation/direct load control versus feedback alone in this study.*

Residential Programs

The most common real-time pricing and data/information program models are those that provide feedback via in-home displays or online portals. Real-time feedback varies in the type of information that is presented to the customer, and the frequency with which it is provided (i.e., hourly, daily). Real-time feedback programs use other behavioral strategies, in addition to feedback. Most frequently, these strategies include goal setting and commitments.

One variable pricing program, SMUD's Smart Pricing Options, was designed to reduce peak demand among its' residential customer base during the summer. SMUD installed in-home displays for customers, and provided rate offers that are combined with feedback tools and educational materials. In addition, customers are rewarded by reduced prices during non-peak periods. Another innovative component of the program was the use of several opt-in and out-opt designs, including: (1) CPP versus TOU rates, (2) opt-in versus opt-out design for both CPP rates and TOU rates, and (3) the offer of an IHD (vs. web-only access to real-time pricing information). A key finding was that opt-in rates were fairly low. Thus, even if average savings from the opt-in participants are higher than average savings from the opt-out participants (since opt-in participants are motivated to save energy due to self-selection), the total savings from the opt-out program is often higher than the total savings from an opt-in program, due to the significantly larger number of participants in the opt-out program (Nexant, 2014).

Commercial Programs

Real-time feedback programs may also be implemented in the commercial sector. However, we did not find any such programs that met our inclusion criteria. Within the commercial sector, real-time feedback is typically not delivered as the primary component of a program, as it is for the residential sector. Instead, real-time feedback is often part of commercial diagnostics programs, where it is bundled with additional information and recommendations, or as part of Continuous Improvement programs, where it is bundled with in-person interactions, training, commitment, goal setting, and follow-through.

5.1.3 Benchmarking Programs

ENERGY STAR Portfolio Manager is a widely-used application that provides commercial and industrial customers with a score that compares a facility or plant to other peer facilities nationwide based upon energy consumption. The program includes goal setting and rewards in the form of recognition from the EPA and other organizations for environmental and operational excellence. In 2012, the Environmental Protection Agency (EPA) reported on buildings benchmarked from 2008-2011 using ENERGY STAR Portfolio Manager, and noted average annual first-year savings of 2.4% (US EPA, 2012). Average percent savings were highest in the retail, office, warehouse and K-12 school sectors, while savings were lower in the grocery, hotel and hospital sectors.

We identified only two evaluations of savings associated with C&I benchmarking programs, both of which were implemented prior to 2010 (2009 and earlier).

Example 1: A 2009 evaluation of National Grid and NSTAR's benchmarking offerings in Massachusetts (which started around 2005) estimated 4-5% net annual electric savings and 3-7% net annual gas savings. About 12-13% of these savings likely came from no-cost/low-cost measures (i.e., not capital outlays). Among savings from measures requiring a capital outlay, between 50-75%

may have come from measures implemented outside of a utility energy efficiency program. Both metrics indicate the potential for "direct" savings from this benchmarking effort. These Massachusetts programs were offered to fairly large customers (demand >200 kW; at least 30,000 square feet). The program included multiple follow-up and continued monitoring (Vaidya, Reynolds, Azulay, Barclay, & Tolkin, 2009).

In the taxonomy, we indicated that Benchmarking programs are primarily for Commercial customers. The only benchmarking programs we could identify with savings are commercial programs. However, a number of benchmarking scores have been developed for the Residential sector, including the Home Energy Rating System (HERS) score, (developed by RESNET), Tennessee Valley Authority's eScore, and the DOE's Home Energy Score (developed by Lawrence Berkeley National Laboratory). Program Administrators typically integrate these scores into existing programs (e.g., audits) rather than developing a stand-alone program around them.

Table 5. First-Year Net Unadjusted Savings Associated with Benchmarking Programs

Program	Behavioral Strategies	Design	Number of Treated Customers	Average Duration for Savings Estimation	Net Unadjusted Electric Savings	Net Unadjusted Gas Savings
ENERGY STAR Portfolio Manager	Feedback; framing; social norms	Opt-in and opt-out (depending on sector)[a]	35,000	1 year	2.4% (Gross; all fuels)	
National Grid / NSTAR Benchmarking [b]	Feedback; framing; social norms	Opt-in	99	1 year	4% - 5% (Net unadjusted)	3% - 7% (Net unadjusted)
NY Benchmarking	Feedback; framing; social norms	Opt-in	428[c]	1 year	1.1%-1.3% (Net unadjusted)	0%-1.9% (Net unadjusted)

[a] Some building and sectors (e.g., public schools) are required to use ENERGY STAR Portfolio Manager. The report does not specify how many participants were opt-in vs. opt-out.

[b] Since these Massachusetts programs used ENERGY STAR Portfolio Manager, they are likely included in the ENERGY STAR Portfolio Manager findings in the row above.

[c] Participants are from K-12 schools or the Commercial Real Estate sectors.

5.2 Diagnostics Programs

Diagnostics programs include residential and commercial audits or energy assessment programs, which may be delivered in-person, online, or with coaching. Typically, these programs record the details of the participants' installed equipment, their building envelope, and in some cases, their operating practices, to develop a list of energy-saving recommendations (using a technician or participant-administered survey). These recommendations may include measure-based actions (identified through building energy modeling), or changes in behavior (operating practices, maintenance, etc.). While audits in and of themselves qualify as a feedback program, we screened existing diagnostic programs to determine whether or not the programs use one or more of the eight social science-based methods to meet our behavioral program criteria.

Program classes included in this description are: (1) Coaching Diagnostics Programs; (2) Online Diagnostics Programs; and (3) In-person Diagnostics Programs.

Within this program category, we found the following:

- The majority of home energy assessment or audit programs did not qualify as "behavioral" programs, per our definition. Based on our review, we were able to identify four residential and two commercial diagnostic programs that fit our definition, which we describe in this section.
- While many diagnostic programs rely on in-person interactions between a property representative and program representative to deliver the program, these interactions are not "peer-to-peer" or with a "trusted community member" in the same manner as social science-based tactics intend (i.e., the interaction is with a designated program technician, not with a peer or trusted community member).
- Though the audit reports and recommendations that customers receive present an opportunity to utilize framing techniques, choice architecture and possibly display social norms, we could not find examples of where this occurred. It is possible that some programs do present recommendations in novel ways, but assessing this would require a comprehensive review of audit/assessment report examples, which are not readily available through public online sources.[14]
- Among the programs that we identified, there were not consistent metrics of success that could be used to make direct comparisons of these programs. As such, we describe several diagnostic programs, but do not attempt to directly compare metrics within those programs.
- There is potential to enhance diagnostics program models with social science techniques. For example, while many diagnostics programs focus purely on cost-saving opportunities, framing techniques could be used to emphasize comfort or home value.

Residential Programs

Despite the lack of social science methods in existing home energy audit and home performance programs, these types of programs are generally effective in spurring behavior change, as evidenced

---

[14] *Franklin Energy teamed with Xcel Energy to test the effectiveness of providing assessment reports that listed best options for saving energy versus reports that did not mention any options for saving energy. Customers were randomized into these two groups. While customers who received the best options recommendation achieved higher mean savings in electricity per customer, these results were not statistically significant (Syring & Laube, 2013). As such, we do not present these in our findings.*

by their widespread use nationwide. Delmas et al. (2013) concluded from a 1975-2011 meta-analysis:

> *"Our results also show that strategies providing individualized audits and consulting are comparatively more effective for conservation behavior than strategies that provide historical, peer comparison energy feedback and pecuniary feedback. This indicates that information delivered in person might be more effective than information provided through other media such as mail or e-mail."*

While many utilities have implemented online home energy analyzers and assessment programs in recent years, we found few with evaluated savings. Instead, it appears that many of these programs are being used to drive customers to other DSM programs (including in-person audits and rebates), and are not responsible for generating direct savings. As such, we briefly describe four residential diagnostics programs that we identified, all of which use non-standard behavioral tactics. Only one of these programs has reported percent savings, therefore, we have not included a table of metrics in this section.

- **"People First" approach of Populus LLC (now CLEAResult**): The program leverages social interactions and feedback behavioral strategies: customers first speak with an Energy Advisor by phone, who assesses the opportunity for savings at a high level, and helps guide customers through the audit and upgrade process. The Energy Advisor is a critical component – helping customers navigate each step of the energy upgrades process, from finding contractors to completing rebate paperwork. Populus was the implementer for the City of Boulder's Better Buildings Neighborhood Program. While we were unable to find an evaluation that reported percent savings associated with this approach, the program achieved 4% participation in EnergySmart[15] program offerings, and 51% conversion to upgrades after a home assessment, compared with 36% conversion for similar BBNP residential programs (Hampton, Hummer, & Wobus, 2012).
- **California Home Energy Survey, offered by Southern California Edison (SCE), Pacific Gas & Electric (PG&E) and San Diego Gas & Electric (SDG&E):**[16] This is a long-standing "self-directed" assessment program that uses feedback as its primary behavioral strategy. First, customers fill out home, equipment and occupancy characteristics through an online, paper (mail-in), telephone, or on-site survey form. Customers are then sent a customized energy assessment and feedback that contains personalized recommendations, energy-savings tips, and information on rebates for which they may qualify. Average annual first-year savings ranged from 0-3.1% for the online survey, from 2.1% for the mail-in survey, 5.6% for the on-site survey, and 6.5% for the telephone survey (Southern California Edison, 2013).
- **Tennessee Valley Authority (TVA) "eScore" scorecard**: The scorecard is used as an in-home energy assessment, and uses feedback and social norms (though these are not delivered in

---

[15] *EnergySmart was designed in collaboration with the City of Boulder Local Environmental Action Division, City of Longmont, and Boulder County Public Health. EnergySmart was funded by the American Recovery and Reinvestment Act through the United States Department of Energy's Better Buildings Neighborhood Program grant, combined with contributions from the City of Boulder's Climate Action Plan tax and the City of Longmont.*

[16] *Note that the California Statewide IOUs have recently adopted a number of behavior change strategies in response to the November 15, 2012 decision mandating that all IOUs reach 5% of customers with behavior change actions. In D.12-11-015, the CPUC states that California IOUs and implementers should focus on using one or more underused behavioral strategies—commitment, feedback, follow through, framing, in-person interactions, energy pricing, rewards or gifts, social norms, and multi-pronged strategies.*

the typical way that social norms are utilized in other programs). The scorecard graphically illustrates a home's current efficiency level. Instead of using a peer comparison or national comparisons, the scorecard compares a customers' current home to their "best home" (a perfect 10 eScore). The scorecard also provides feedback, by highlighting and encouraging the simplest and highest-value energy actions each household could take. This program is fairly new and has not been evaluated yet. Other scorecard programs, for example, a scorecard co-developed by the Earth Advantage Institute and Energy Trust of Oregon, also graphically illustrate a home's energy score, based on total energy use, and provide recommended actions. Similarly, this scorecard has not been evaluated.

- **BC Hydro Online Diagnostics Program:** This long-running online diagnostics program incorporates goal-setting, online tracking, follow-through, and rewards (e.g., $75 reward for reaching 10% savings). Though evaluations do not report percent savings, first-year savings are between approximately 450-490 kWh per household, and second and third-year savings (for households who continue to participate) range from 200-300 kWh per household. For customers who decided not to participate in later years, savings do not appear to persist (BC Hydro, 2014).

Commercial Programs

On-site commercial energy assessment programs typically feature direct install of low-cost measures and recommendations for follow-up measures. We did not find examples of "basic" energy assessment programs that use behavioral intervention tactics and have evaluated savings.[17] While we discuss growth and trends in program offerings below, we do not attempt to directly compare programs. Instead, we briefly describe two commercial diagnostics programs that we identified, both of which use non-standard behavioral tactics.

- **Commercial and Industrial (C&I) Remote Audits**: First Fuel offers remote audits to C&I customers, which includes two "levels" of service. First, utilities and the building manager can use the FirstScreen tool to screen hundreds or thousands of buildings for retrofit and operational savings potential. Then, the FirstAudit tool can be used to dive deeper into each building's performance. This tool provides feedback with customized, actionable, and building-specific recommendations for retrofit and operational savings. The New Buildings Institute offers a similar tool, called FirstView. This software providers building owners with more targeted feedback and energy performance information by using monthly utility bill data and building characteristics. Pulse Energy (now part of EnerNOC) and PG&E (Gridium) also provide remote audits. However, we could not find program evaluations with reported savings for any of these programs.
- **Business Energy Analyzers**: Commonwealth Edison's (ComEd) offers a Business Energy Analyzer, implemented by Agentis, which graphs consumption information, provides personalized recommendations (typically framed as cost savings), identifies operational savings potential by time of day, and provides some benchmarking analysis and/or peer comparison. The recommendations frame opportunity in terms of savings and payback, and customers can track online whether they've completed each recommendation. The program was piloted as an opt-out model using a series of mailers with a more basic report and more

[17] Programs labeled as "strategic energy management", "continuous energy improvement" and "benchmarking" could all be considered diagnostic programs. Since these programs are multi-pronged with multiple tactics, they appear in other areas of the taxonomy.

detailed information on the online user interface. It is now offered to all of ComEd's C&I customers. However, only the pilot has been evaluated. For the 6,000 customers who were sent mailers and encouraged to use the website, savings were about 0.2% (Navigant, 2014a). Savings among those who use the online energy analyzer has not been evaluated.

6.0 Social Interactions Family Findings

We present our findings for the social interactions family of programs in this section, defined as programs that share information through social interactions. These interactions may occur either online or in-person. Whereas calculus programs rely on rationality, programs in this family feature "sociability and belonged experience" (Mazur-Stommen & Farley, 2010). This family includes the following program categories, listed in order of the prevalence of evaluated programs: (1) Community-Based Programs, and (2) Competitions.

6.1 Community-Based Programs

Community-based programs are defined as those programs that "take advantage of some form of face-to-face social interaction" (Mazur-Stommen & Farley, 2010). Community-based programs operate under the assumption that people are more likely to trust information that they receive through in-person interactions; for example, from a trusted community member, from other individuals who are similar to themselves (i.e., a neighbor), or at community events. Utility companies have implemented community-based programs for several years, and continue to do so today. Recently, the United States Department of Energy (DOE) funded community-based programs through the American Recovery and Reinvestment Act (ARRA), resulting in an increase in these types of programs. ARRA provided funding to 41 grantees as part of the Better Buildings Neighborhood Program (BBNP). Many grantees were state or local governments, though they typically partnered with utilities to administer community-based programs. Among these programs, multi-channel outreach, in-person interactions, such as community sweeps, and competitions were frequently used.

Within this program category, we found the following:

- We reviewed five residential and four commercial community-based programs,[18] implemented between 2010-2012, that met our criteria of having both behavioral interventions and evaluated savings.
- When reporting savings, community-based efforts report gross savings, due in part because these programs often operate to channel participants to other programs.
- Of the programs reviewed, Community-based program evaluations report from 12% to 30% per premise electric savings for the residential sector, and from 10% to 31% per premise electric savings for the commercial sector.
    - These savings are derived from energy upgrade components of these programs.
- Of those reviewed, community-based programs are most often paired with the following social science interventions, listed in order of their prevalence in programs: (1) in-person interactions (such as neighborhood sweeps or door-to-door outreach); (2) outreach from trusted community leaders (such as a town Mayor); (3) feedback (including real-time); and (3) competitions.
- Evaluation methods used for community-based programs and programs with peer-to-peer components include pre/post billing analyses using matched comparison communities and

[18] *Note, the ILLUME team reviewed seven other programs, which used behavioral strategies, including the Otter Tail Campus Energy Challenge. However, these programs did not have evaluated savings and were not included in this formal count.*

development of gross savings to estimate home or business energy upgrade components of these programs.

We report on community-based programs delivered in Arizona, Michigan, Washington, Colorado, and Rhode Island. Table 6, below, summarizes all key benchmarked metrics for community-based programs included in this study. Following Table 6, we describe residential and commercial community-based programs.

Table 6. Savings Results for Community-Based Programs[a]

Program Name	Behavioral Strategies	Average Duration for Savings Estimation	Residential Participation Rate	Residential n	Residential Electric Savings	Residential Gas Savings	Commercial Participation Rate	Commercial n	Commercial Electric Savings	Commercial Gas Savings
Energize Phoenix[b]	Social interactions: Door-to-door outreach, events	1 year[c]	NR	2,014	12%[f]	NA	NR	375	10-17%[f]	NA
Michigan Saves[b]	Social interactions: Targeted neighborhood "sweeps" in 58 locations	1 year[c]	14%	7,689	14%[e, f]		NR	81	31[f]	
Seattle Community Power Works[b]	Social interactions: events, phone center support, contractor training	1 year[c]	NR	3,070	30%[f]		NR	153	13-18%[f]	
RePower Bainbridge Island Energy Upgrades[b]	Framing; Feedback (community-level) peer-to-peer interactions	1 year[c]	NR	977[d]	30%[f, g]		NR	238[d]	NR	NR
Energy Management Teams – Coordinator Resource Pilot	Social interactions; goal-setting	NR	NA	NA	NA		NR	5	NR	8.1%
Otter Tail Power On Community Energy Challenge Program	Social interactions; goal-setting; education and training	1.5 years	NR	205	234,725[h]		NR	10	156,995[i]	NR

[a] All of the programs in this table were an opt-in design, therefore we did not include a column for opt-in vs. opt-out design in this table.

[b] Results from these programs were not evaluated by a third-party, and reflect savings from energy upgrades, and were typically calculated using deemed savings or modeled savings (from upgrade software).

[c] For all programs in the table except BC Hydro Workforce Conservation, the savings displayed in the table are savings associated with home energy upgrades or business energy upgrades, typically delivered through existing programs. As such, the savings estimates are based on first-year program-reported measure savings, typically estimated from building energy models (implementer software) or deemed savings.

[d] Participation counts are for the home energy upgrades and business energy upgrades components of the program. An unknown number of customers may have participated in other components, such as direct installs and energy conservation.

[e] Home and business energy upgrades in Michigan Saves included electric and gas measures. The final report did not specify whether percent savings results applied to electricity, natural gas, or both fuel types.

[f] Seattle Community PowerWorks calculated percent savings from home and business energy upgrades as MMBtu savings across all fuels as a percent of baseline consumption. Separate savings values for electric and gas measures are not available.

[g] Home energy upgrades in RePower Bainbridge and Bremerton included electric measures and weatherization/heating measures that were heating fuel-agnostic. The final report did not specify what fuel percent savings apply to (possibly all fuels, including oil).

[f] Savings are gross savings.

[h] This figures includes savings generated from the behavior components of the program. Equipment-based savings were also generated, but are excluded in this figure. The program reports only community-wide electricity savings, not percentage savings.

[I] Commercial savings include savings from the Commercial Behavioral Change and School ReDirect programs. The program reports only community-wide electricity savings, not percentage savings.

Residential Programs

The most common residential community-based program models are those that use in-person interactions, including neighborhood sweeps, community events, and door-to-door outreach within a community. As noted above, many of these programs include peer-to-peer engagement as a specific outreach strategy.

At the national level, a preliminary impact evaluation was conducted across multiple DOE grantees. Among four grantees for which they could gather sufficient billing analysis data for residential retrofit participants, they found gross electric savings of 12% and gross gas savings of 11.1% from upgrades, representing a 79% realization rate in gross savings from what was reported in each project's preliminary report to the DOE. They also estimated a net-to-gross ratio of 83%, bringing electric and gas savings to 10% and 9.2% on average, respectively (Research into Action, 2013).

A notable community-based program is RePower Bainbridge's "Island Energy Dashboard", which shows system-wide peak demand at the community level. The program implementer installed dashboards at local businesses all over the island (via tablets). The graphic's simplicity and use of common, everyday symbols is an example of the use of non-price "framing". Additionally, displaying data at the community level, rather than at the household level, encourages thinking about energy as a collective issue, and provides a sense of urgency when the needle is in the red or orange. The program evaluation did not report savings or level of influence specific to the Island Energy Dashboard (Kraus, 2014). However, savings were reported for the existing Home Energy Upgrades program, which is part of this program effort. Among 977 participants, 30% electricity gross savings per premise was reported. It should be noted that this estimate may be inflated due to evaluation design issues.

As another example, National Grid has implemented two community-based programs in Rhode Island, both with a goal of reducing energy consumption and peak summer demand. The first program, Energy Action Aquidneck and Jamestown, was an 18-month effort with enhanced program marketing, engagement with community leaders, community outreach and local events (targeted to both residential and commercial customers). The second program, the System Reliability Procurement pilot, is a multi-year effort aimed at reducing demand in a capacity-constrained areas among the residential sector. Neither program measured average percent savings per customer, but both programs reported savings that were "incremental" to business-as-usual savings in audit and rebate programs (for residential and commercial customers) – i.e., savings that likely would not have been achieved without the community-based efforts. The first pilot (Energy Action) achieved 13% incremental electric savings and 15% incremental gas savings in Residential programs, and 53% incremental electric savings in Commercial programs (Opinion Dynamics, 2011). The second pilot (System Reliability Procurement) achieved 53% incremental electric savings in National Grid's residential retrofit program in the first 18 months of the pilot (Opinion Dynamics, 2014b).

Some residential community-based programs also track increases in program participation. Michigan Saves, which conducted 58 sweeps in targeted geographic areas, achieved a 14% participation rate (on average) for their base package of services. *The Rhode Island Energy Challenge: Find Your Four* was a community-based effort launched in 2013. It has not yet been evaluated for energy savings, but did report a 5% household participation rate. The premise of the program was to get households to take just four energy-savings actions in their homes (NY Customer Engagement Committee, 2014).

It is worth noting that we found at least one feedback program (Tucson Electric Power's (TEP) PowerPartners) that tried to encourage peer-to-peer interaction through an online forum, which included an "energy expert" moderator weighing into the forum. Utilities are increasingly attempting to stimulate peer-to-peer interaction in social media (SMUD's SmartPricing Options; Clark Public Utilities' "Social Energy" campaign). However, we did not find any savings estimates associated with the peer-to-peer component of these programs.

Commercial Programs

We identified four commercial community-based programs that met our selection criteria. These include programs in Arizona, Michigan, and Washington, all of which are highlighted in Table 6, above. Similar to the residential sector, these programs use in-person outreach and trusted community members to prompt energy saving actions, in addition to personalized feedback.

As an example, BC Hydro's Workplace Conservation Awareness program is centered around peer-to-peer interactions with energy champions, who are responsible for planning the program, setting goals, engaging their colleagues with behavior change campaigns, and providing feedback on energy performance. In many sectors, net program savings were not statistically significant in the first or second program years, but in the second program year, net electric savings for K-12 schools were 3% of baseline consumption. The lack of statistically significant savings may be related to a small sample size of 300 buildings participating, across multiple sectors (BC Hydro, 2015).

Another example is that of a pilot program undertaken in Minnesota by Franklin Energy and the Minnesota Energy Resources Corporation. The pilot evaluated the impact of providing a dedicated energy management team coordinator to large commercial and industrial customers. Annual gas savings achieved by participants was 8.1%, as compared to 6.4% among the comparison group, which did not receive support from an energy management coordinator. The program did not report differences in electric savings for participants and the comparison group (Syring, Gorell, & Laube, 2013).

Also in Minnesota, Otter Tail Power Company launched a community-based program implemented for both residential and commercial sectors. The program leveraged in-person interactions, goal-setting, feedback, and education and training. Community-wide total electric savings for the residential sector were reported as 234,725 kW, while community-wide total commercial electric savings for the commercial sector were reported as 156,995 kW (Otter Tail Power Company, n.d.).

While we did not find independent third-party evaluated savings, it was worth noting AEP Ohio's Continuous Energy Improvement program. The program works with commercial customers to develop energy teams. These teams engage employees, identify opportunities for improvements, and track progress toward goals. In addition, Minnesota Power has recently formed community energy groups for business customers. The group has energy coordinators/managers from various businesses that meet periodically with representatives from Minnesota Power to share ideas for energy efficiency projects. This program does not yet have evaluated results, and hence, is not included in our benchmarking analysis.

6.2 Competitions

Competitions include programs in which individuals or organizations compete in events, contests, or challenges. As described by Jones and Vine (2015), "competitions provide a set of rules, mechanisms to track results, and recognition to participants for their progress at achieving a specified objective." Originally, our taxonomy included both peer-to-peer competitions and inter-personal competitions as program classes within the Games program category. However, upon completing our research, we found that Competitions is a more appropriate program category, and as such, have renamed this category. Within the Competition program category, our benchmarking analysis compares the following program classes, in order of evaluated data: (1) Group competitions that feature games (called Game competitions), and (2) Group competitions that do not specifically feature games or game activities (called non-game competitions).

Competitions operate under the assumption that individuals or teams will be motivated to compete against one another. Competitions are a component of games, and may also include behavioral strategies, such as social norms, feedback, goal-setting, and commitments. Not all games are competitions, though games frequently incorporate competitions as a behavioral strategy. While there is no academic definition of games, gamification refers to the "incorporation of game design elements or strategies into real world applications" (Jones & Vine, 2015). Other common elements of games include incentives and rewards. It should be noted that programs operating within this category are relatively new, with short program durations. Thus, results should be interpreted with caution.

Within this program category, we found the following:

- Across these program classes, we reviewed 11 residential and eight commercial competition programs, implemented during 2011-2014 that met our criteria of having both behavioral interventions and savings.
    - Of these, few game competition programs have been evaluated for energy savings by a third-party evaluator.
    - There are very few competition programs that do not feature game elements and which also have evaluated savings (n=3).
- To date, few competition programs that feature games have been implemented beyond the pilot stage.
    - Of these, program duration is typically less than six months.
- Very few game competitions have undergone rigorous evaluations, resulting in a relative dearth of data related to savings metrics. Of those evaluated, net unadjusted energy savings are typically less than 5%.
- Of those reviewed, competition programs are most often paired with the following social science interventions, listed in order of their prevalence in programs: (1) social norms; (2) feedback (which may be real-time); (3) peer-to-peer interactions; (4) goal-setting and public commitments; and (5) rewards and incentives.

6.2.1 Game Competition Programs

Group competitions are defined as those in which groups of players (i.e., neighborhoods, communities, or organizations) compete against one another. These competitions may be inter-group competitions (when different communities compete against one another) or intra-group competitions (when organizations in a similar community compete against one another). In this program class, we report specifically on competitions that feature games or game activities. Of the evaluations we reviewed, the most common behavioral strategies include competitions, peer-to-peer interactions, comparative feedback, goal-setting and public commitments, and prizes and rewards.

Residential game competitions generate net unadjusted electric savings ranging from 3-14%, while commercial group game competitions generate net unadjusted electric savings ranging from 5-21%. Most game competitions have not targeted, or have not reported, gas savings, and results have not been verified by third-party evaluators. It is important to note that, even among those games that report savings, these metrics are not directly comparable. This is due to differences in baseline definitions, how percentage savings are reported (i.e., relative to baseline use among participants or relative to control group participants), evaluation methodologies utilized (i.e., billing analysis, interview method to estimate savings associated with specific actions), and the time periods of game competitions (ranging from two weeks to 12 months). Persistence has been rarely measured (Jones & Vine, 2015). Further, a 2015 literature review collected data on approximately 40 games, and presented case studies of 22 of those games that either could, or currently are, being utilized by utilities. That review noted that, "more evaluation is needed, including evaluations of first-year savings and persistence" (Grossberg, Wolfson, Mazur-Stommen, Farley, & Nadel, 2015).

We reviewed 11 residential and five commercial group game competitions,[19] which met our criteria for use of behavioral strategies, and which had evaluated savings data. Some of these competitions featured both group competitions, and individual competitions, and as such, we include these in this section. One competition (Biggest Energy Saver) only featured competitions between households; however, we also include that program in this section. We did not identify other programs that featured competitions focused solely on peer-to-peer competitions. The tables below summarize all key benchmarked metrics for group competitions that feature games.

[19] Note, the ILLUME team reviewed nine other programs; however, these did not have evaluated savings and were not included in this formal count.

Table 7. Savings Associated with Competition Game Programs for Residential Sector

Program	Behavioral Strategies	Design	Average Duration for Savings Estimation	Participation Rate	Participants (n)	Electric Savings	Gas Savings
Cool California Challenge	Community competition; social norms (comparative feedback); peer-to-peer interactions; rewards	Opt-in	5 mos.	NR	2,700 households	14% (Gross)	0%
Energy Smackdown	Community competition; social norms (comparative feedback); peer-to-peer interactions; rewards	Opt-in	12 mos.	NR	100 households (3 communities)	14% (Gross)	17% (all heating fuels)
Kansas Take Charge Challenge	Community competition; commitments/goal-setting peer-to-peer interactions; rewards	Opt-in	9 mos. (in year 2)	NR	100,000 households (6 communities) (in year 2)	5% (for top-performing community; gross)	NA
Western Mass Saves Challenge	Community competition; goal-setting; feedback via online portal; rewards	Pre-selected (by program)	8 mos.	NR	2,000 households (4 communities)	0.1%-2.3% (Per community among participants; gross)	NA
SDG&E Energy Challenge (California)	Household competition on behalf of schools; social norms; peer-to-peer interactions; rewards	Opt-out weekly emails; opt-in online platform	9 mos.	27% became "engage customers"	5,634 competing households	6% summer and 2% winter (net unadjusted); 2.2% on-peak demand reduction	NA
Biggest Energy Saver (part of SDG&E Energy Challenge)	Household competition; real-time feedback via IHD; rewards	Opt-out	2 mos.	NA	200 households	11% (Gross)	NA
Chicago Neighborhood Energy Challenge	Individual and building competition; feedback; peer-to-peer interactions; rewards	Opt-in with pre-selected buildings	6 mos.	80%	600 residents / 7 multi-family and senior housing buildings	5% (Gross)	10% (Gross)

Program	Behavioral Strategies	Design	Average Duration for Savings Estimation	Participation Rate	Participants (n)	Electric Savings	Gas Savings
Southern Maryland Electric Cooperative Energy Savings Challenge	Individual and team competition; goal-setting; feedback; rewards	Opt-in	3 mos.	NR	201	NR[b]	NA
Reduce the Use in District 39 (New York)	Monthly feedback with energy-saving tips; social norms	Opt-in	12 mos.	NR	161 households	4% (Gross)	NA
iChoose (Miron Construction, Wisconsin)	Individual and team competitions; points; rewards	Opt-in	6 mos.	67%	220 employees	4% (Gross)	~0%

[a] It is unclear whether these savings are gross or net. While the program design did include matched comparison communities for use in evaluation, we do not know if average savings per participant calculations in challenge communities leveraged the matches.

[b] Though overall average percent savings were not reported, 38% of participants achieved the 3% reduction goal (gross).

Table 8. First Year Net Unadjusted Savings Associated with Competition Game Programs in the Commercial Sector

Program Name	Behavioral Strategies	Design	Average Duration for Savings Estimation	Participation Rate	Participant n	Electric Savings	Gas Savings
iChoose (Milwaukee Fire Department, Wisconsin)	Team competition	Opt-in	2 mos.	14% of employees	29 buildings / 130 participants	6.6% (Net unadjusted)	NA
Kukui Cup (Hawaii)[a]	Dormitory competition; points; rewards	Opt-in	2 wks. (year 3)	33% of dorm residents	350 students	8% (Gross)[a]	NA
Campus Conservation Nationals	Dormitory competition; goal-setting; feedback; rewards	Opt-in for school; opt-out for students	3 wks.	NA	200,000 students (year 2)	3-4% (Gross)	NA
Carbon4Square (Oregon)	Competition between buildings; Feedback	Opt-in	NR	NR	85 buildings	NR	NR
Duke Smart Energy Now	Real-time feedback via lobby kiosks; energy champions (peer-to-peer), training, pre-packaged "campaign"; behavioral experts	Opt-in	NR	NR	59 buildings	6.9% (Net Unadjusted)	NA
BC Hydro Workforce Conservation	Energy champions (peer-to-peer); real-time data/feedback; commitment; rewards	Opt-in	1 year	NR	300 sites	0-3% (Net unadjusted)	NA
SnoPUD Behavior-Based Energy Efficiency Pilot	Within-store competition; real-time feedback via in-store displays; education	Pre-selected	1 month	NA	10 stores	2% (Net unadjusted)	NA
Boulder 10 for Change Challenge	Business-to-business competition; peer-to-peer interaction; commitments; goal-setting	Opt-in	1 year	NR	100 businesses	4-8% (Gross) [b]	NA

El Paso Energy Savings Challenge	Competition between buildings; commitment; goal-setting	Pre-selected	6 mos.	NR	Year 1: 12 libraries Year 2: 34 fire stations	Year 1: 262,000 kWh Year 2: 100,000 kWh[c]	NA

[a] The authors of the Kukui Cup Challenge reported several limitations of this energy savings data, including baseline values that may be "grossly inaccurate", inflated savings, and differences in teams competing in the challenge (Johnson et al., 2013).

[b] Participants in the Boulder 10 for Change Challenge saved electricity and natural gas. The program report does not specify whether the 8% savings result is electric or gas; it may be across both fuels.

[c] El Paso Energy Savings Challenge electric savings are reported as total kilowatt hours saved.

Residential Programs

The most common residential game competition programs include those in which residential customers, or students residing at universities, compete on teams for the greatest reductions in energy savings. Residential game competition programs include several components beyond the competition itself. These include feedback, which is sometimes real-time, comparisons with other "players" of the game, setting goals, making public commitments, and prizes and rewards for achieving specific goals.

As an example of a university program, the Kukui Cup was a competition that included real-world and online activities for students residing in dormitories at the University of Hawaii. The competition was held during the fall 2011 semester, and included participating teams of 50 students each. Students made public pledges and received points for adopting specific energy-saving behaviors. While several teams appeared to achieve a 10-16% reduction in their energy usage during the intervention period, study authors concluded that, "the almost exclusive focus…on measuring percentage reductions from baseline during competitions leads to (1) invalid claims regarding savings and (2) an incentive to engage in unsustainable behaviors" (Johnson et al., 2013).

As an example of an intra-community competition, in 2013, San Diego Gas and Electric reported on its Biggest Energy Saver program – a community-based challenge in which customers earn prizes, points, and rewards for certain behaviors. Customers "played" the game individually and also for one of 39 participating San Diego schools. On-peak demand reduction of 2.2% was achieved among participants in the challenge (Braithwait, Hansen, & Hilbrink, 2013). SDG&E also reported overall pre premise net energy savings of 2% during the summer and 6% in the winter (Reguly, 2013).

Commercial Programs

Game competitions are increasingly being implemented within the commercial sector, particularly for building-to-building competitions. Similar to the residential sector, these programs frequently include several behavior strategies in addition to competitions. These include peer-to-peer interactions, often with a "messenger", commitments, goal-setting, and prizes.

As an example, Cool Choices designs and implements game competitions. Their first game, called iChoose, was implemented from May through November, 2011 at Mirion Construction. A total of 220 employees participated, and competed individually, and in teams, to earn points for specific behaviors taken to reduce energy consumption. The Energy Center of Wisconsin conducted a billing analysis, comparing pre- and post-intervention consumption for electricity and gas, for 70 participating employees. The billing analysis results demonstrated an average of 4% savings per participant (Bensch, 2013). However, these numbers should be interpreted cautiously, given an extremely small sample size and lack of control group. While persistence was measured, it was done so using an interview method to approximate energy savings associated with specific actions. Billing analysis was not completed to measure persistence. A similar game was implemented by Cool Choices at the Milwaukee Fire Department. Participating firehouses achieved a net 6.6% reduction in energy consumption, as compared to non-participant firehouses. Persistence was not measured.

6.2.2 Non-Game Competitions (Commercial Only)

Non-game competitions are defined as those that programs that use competition elements, but that do not specifically develop games or game activities as part of the program. Instead, these programs feature competitions, coupled with real-time feedback, and "energy champions" who promote the program and encourage participation. Non-game competitions generate net unadjusted savings ranging from 0%-6.9%. We identified three non-game competitions for the commercial sector. We did not identify any non-game competitions for the residential sector. Table 9, below summarizes all key benchmarked metrics for non-game competitions within the commercial sector.

Table 9. First-Year Net Unadjusted Savings Associated with Commercial Competition Non- Game Programs for the Commercial Sector

Program Name	Behavioral Strategies	Design	Average Duration for Savings Estimation	Participation Rate	Participant n	Electric Savings	Gas Savings
Duke Smart Energy Now	Real-time feedback via lobby kiosks; energy champions (peer-to-peer), training, pre-packaged "campaign"; behavioral experts	Opt-in	NR	NR	59 buildings	6.9% (Net Unadjusted)	NA
BC Hydro Workforce Conservation	Energy champions (peer-to-peer); real-time data/feedback; commitment; rewards	Opt-in	1 year	NR	300 sites	0-3% (Net unadjusted)	NA
SnoPUD Behavior-Based Energy Efficiency Pilot	Within-store competition; real-time feedback via in-store displays; education	Pre-selected	1 month	NA	10 stores	2% (Net unadjusted)	NA

As an example of a commercial non-game competition, in 2011, Duke Energy developed a program for commercial office buildings in downtown Charlotte called “Smart Energy Now”. Occupants saw real-time feedback on kiosks in building lobbies; there was web-based feedback; and energy champions sent emails, newsletters or mail about building energy efficiency progress. The program engaged facility managers and executives to buy into the effort, and behavioral experts led initiatives to engage occupants. Energy champions were a major component of the program, as the champions chose which of the pre-developed engagement campaigns (developed by Duke) to deploy, and how to keep their colleagues engaged. As such, the program had a strong peer-to-peer component. Throughout 2012-2013, 59 office buildings with 70 unique Duke accounts competed in the challenge, and billing analyses showed average net unadjusted savings of 6.9%, and net adjusted savings of 6.2% (TecMarket Works, 2014).

7.0 Cognition Family Findings

Cognition programs are defined as programs that appeal to emotions and/or rely primarily on delivering information to customers as a means of driving change. These programs provide general information and differ from the Calculus family, which provides highly customized and targeted information focused on direct behavioral feedback.

This family includes only one program category: Education and Training. Most Cognition programs do not get to savings, but rather, focus on raising awareness generally.[20]

7.1 Education and Training Programs

Education and training programs are defined as those programs that provide general education to promote the adoption of energy efficient behaviors, or to encourage building changes that will result in energy savings. Within the Education and Training program category, our benchmarking analysis compares the following program classes, in order of evaluated data: (1) Continuous Improvement programs; and (2) K-12. The former program class is only relevant to the commercial sector, and as such, we only present commercial findings for this class. Originally, our taxonomy included a third program class: Building Operator Certification programs. However, the Continuous Improvement programs that we identified include training programs, though these are not called certification programs. As such, we have omitted Building Operator Certification programs as its own class.

Education and Training programs were only included in our benchmarking analysis if they specifically used social science-driven behavioral strategies, and had evaluated savings. Many programs that fall within this program class did not meet these two criteria, and as such, are not included in our benchmarking analysis. Of those programs that did use behavioral strategies, most frequently, these programs rely upon training and information, goal setting, commitment, in-person interactions, and feedback. Below, we present limited findings for this program class.

Within this program category, we found the following:

- We reviewed five commercial CIE programs, implemented between 2009-2014, which met our criteria of having both behavioral interventions and evaluated savings.
- When reporting savings, CIE programs report gross savings, due in part to challenges in measuring net savings associated with this class of programs.
- Of the CIE programs reviewed, program evaluations report from 0% to 22% gross per premise electric savings, and from 0% to 23% gross per premise gas savings.
    - Two CIE programs reported gross adjusted savings, attempting to exclude savings from measures installed during the program period.
- The most common evaluation approaches for CIE programs include pre-post billing analyses. CIE program evaluations did not use comparison group analyses.

---

[20] *Originally, our taxonomy included the Market Actor program category within the Cognition family. Market Actor programs include training courses for builders, and trainings for home contractors on energy efficiency equipment, installation, and maintenance practices. The program classes defined in our program taxonomy focus on encouraging behavior change among end users. Market actors are often involved in those programs (for example, market actors are often an important part of home performance programs), but behavioral strategies typically target the end consumer of energy. As such, we do not report specifically on market actor programs, as the efforts of market actors are captured in other program classes.*

- We reviewed two K-12 programs, implemented between 2006-2013: both of these programs used behavioral strategies, but only one had evaluated savings.
    - Many K-12 programs do not have evaluated savings, and as such, are not included in this analysis.
    - The one K-12 program with evaluated savings did not describe their evaluation approach.
- Of all Education and Training programs (including CIE and K-12 programs) reviewed, the following social science interventions are used frequently, listed in order of their prevalence in programs: (1) training; (2) goal-setting; (3) commitment; (4) in-person interactions; and (5) feedback.

7.1.1 Commercial Continuous Energy Improvement and Strategic Energy Management Programs

Continuous Energy Improvement (CEI) programs are defined as programs that present commercial building energy information to building managers and operators in an actionable way. While such programs incorporate aspects of calculus (i.e., feedback), these programs typically also leverage in-depth educational engagement, hands-on training of facility managers and building operators, often via an energy champion, and goal-setting and commitment. Program staff typically works with the facility manager and operators over time to review usage information, and advise or help implement building changes. CEI programs have been most broadly implemented within the Pacific Northwest region of the United States. According to the Northwest Energy Efficiency Alliance, program models must incorporate: (1) goal setting; (2) executive commitment (for example, dedication by senior management or energy champions); (3) appropriate training to help achieve set goals; and (4) tracking of progress toward the goal (DNV KEMA and Research into Action, 2014).

CEI programs generate gross per premise savings ranging from 0% to 22% for electric and from 0% to 23% for gas. While a handful of Program Administrators, particularly in the Pacific Northwest and British Columbia, have been offering these programs for years, they are generally not widespread, nor evaluated for energy savings.

We reviewed five different CEI programs, implemented in British Columbia, California, Washington, and Oregon, which met our criteria for use of behavioral strategies, and which had evaluated savings data. Among these programs, customized training, goal-setting, commitment, in-person interactions, and feedback are frequently used. Table 10 below, summarizes all key benchmarked metrics for CEI programs included in this study. Following Table 10, we describe CEI programs.

Table 10. Savings Results for Commercial Continuous Energy Improvement Programs

Program	Behavioral Strategies	Design	Participation Rate	Participants (n)	Average Duration for Savings Estimation	Electric Savings	Gas Savings
ETO Strategic Energy Management	Workshops and energy assessment, track performance, executive sponsor, energy champion, peer-to-peer networking	Opt-in	NR	12	1 year	4.7%-6.7% (Adjusted Gross)[a]	3.8%-9.8% (Adjusted Gross)
BPA Energy Management Pilot	Technical assistance and training, co-fund staff time for O&M, track and tune projects	Opt-in	NR	16 electric / 2 gas	1 year	2.7% (Adjusted Gross)[a]	25% (Adjusted Gross)
BC Hydro Continuous Optimization	In-person interactions, training, feedback	Opt-in	NR	115	1 year	7% (Gross)[b]	11% (Gross)
CPUC Continuous Energy Improvement	Energy assessment, technical assistance, management plan, organization assessment and commitment from senior management	Opt-in	NR	3	1.5 years	-2%-5.2% (Gross)	2%-18% (Gross)
Puget Sound Energy Resource Conservation Management	Incentives, dedicated staff, goal-setting, Resource Conservation Manager	Opt-in	NR	864	1 year	1-22% (Gross)	0-23% (Gross)

[a] ETO and BPA savings are gross savings specific to O&M actions – they exclude savings from measures installed during the program period

[b] The program also achieved an 11% reduction in peak demand savings during the analysis period.

The most common CEI program models include those that use goal-setting, commitment, in-person interactions (for example, with an energy champion), customized training, and feedback. A 2012 Consortium for Energy Efficiency (CEE) study analyzed savings from NEEA, Bonneville Power Association (BPA), and the few California CEI programs conducted through that date, and concluded, "In general, facilities participating in strategic energy management programs realize approximately 3% electric savings from O&M measures during the year which the programs were evaluated" (CEE, 2012).

Among all five programs we reviewed (Table 10, above), savings were evaluated with site-specific pre-post billing analysis using regression models. These analyses resulted in gross savings for electric and gas, ranging from 0%-7% of baseline consumption, and from 2%-25% of baseline consumption, respectively (Cadmus, 2013a; ETO and Cadmus, 2014; Cadmus, 2013b; IndEco 2011). Only one program administrator, the Energy Trust of Oregon (ETO), attempted to isolate operational and maintenance (O&M) savings from physical and measure-based savings. These savings included 4.7%-6.7% gross electric O&M savings, and 3.8-9.8% gross gas O&M savings.

Other Program Administrators offer CEI programs, but we could only find evaluated results for these programs. For example, Vermont launched its' CEI Pilot in 2014, but the program has not reported savings results.

7.1.2 K-12 and Campus Programs

K-12 programs are defined as classroom-focused educational efforts that aim to generate savings within school facilities. Relatedly, campus-based educational efforts may target a variety of behaviors and end-users, including administrators, faculty members, and students, to change energy-using practices at various levels of the organization. Originally, K-12 and Campus programs were defined as separate program classes within our taxonomy. Given limited program findings, we have combined K-12 and Campus program classes into one section. We did identify two campus-based programs which included educational efforts; however, those programs models focused primarily on competitions, and as such, those programs are reported on in the Competitions section of this report. Of the evaluations we reviewed, we identified the following behavioral strategies: education, feedback, and rewards.

Within this program category, we found the following:

- There is limited savings-related data on behavioral programs implemented in the K-12 settings.
- We identified two programs, implemented from 2006-2013, that met our criteria of having behavioral interventions. Only one of these programs has been evaluated for energy savings.

Given limited findings for this program class, we do not present a table of benchmarked metrics for these two programs. Instead, we describe both programs, below.

PowerSave Schools is a program model developed by the Alliance to Save Energy. The program provides tools and curricula for engaging students with energy data, including getting students involved in collecting and analyzing the data. During the first seven years of operation (through 2013), 600 schools worked with the Alliance to Save Energy to implement the PowerSave Schools program. Gross per premise electric savings among California schools averaged 12% over eight months of the school year (Alliance to Save Energy, 2012).

In 2014, New Jersey launched a "Conserve to Preserve" community rewards program that gives money to schools when households complete online energy audits (New Jersey Natural Gas, 2014). The program uses both education and reward behavioral strategies. The program has not yet been evaluated.

Section 2: Evaluation, Persistence, and Measure Life Overview

Section Two of this report details currently-used and recommended evaluation approaches for CIP programs, and discusses policy issues around claiming savings for CIP programs, including persistence and measure life.

8.0 Recommended Evaluation Approaches

In this section, we: (1) briefly outline recommended evaluation approaches utilized for estimating savings for behavioral programs, (2) provide an overview of the evaluation methods currently used in behavioral CIP programs to demonstrate the level of rigor with which each behavioral CIP class has been evaluated to date, and (3) provide a *decision-making* focused framework for selecting between evaluation methods for various behavioral CIP designs.

In most cases, behavioral CIP efforts aim to measure a percent savings reduction per premise engaged in the program using billing or meter data. For example, savings may be expressed as a two percent reduction in usage per participating household. The behavioral CIP efforts that have been successful in claiming energy savings have used either experimental or quasi-experimental designs. This is the evaluation precedent set for behavioral program evaluations. However, depending on the program design, evaluator will have varying levels of success implementing these methodologies.

Here, we briefly discussion methodologies that have been used to measure savings on a per-premise basis. These methods are divided into three groups: (1) experimental designs, (2) quasi-experimental designs, and (3) preponderance of evidence-based approaches. Several studies have outlined the specific impact evaluation approaches that may be used for experimental design (EPRI, 2010; Vine, Sullivan, Lutzenhiser, Blumstein, & Miller, 2014), behavioral feedback programs (SEE Action, 2012) as well as for rate- and event-based programs (Cappers, Todd, Perry, Neenan, & Boisvert, 2013). Our work does not reiterate the technical application of these methods nor the best practices associated with conducting a technical evaluation.

8.1 Establishing the Counterfactual

The primary objective of impact evaluation is to (a) measure changes in energy consumption during the program period among customers who received or were exposed to the program, and (b) provide sufficient evidence to demonstrate that a program caused these changes in energy usage. In the case of behavioral programs, published best practices have largely focused on experimental design (Todd et al., 2012, Cappers et al., 2013, Vine et al., 2014) to demonstrate program effects. However, significant advances have been made in applied evaluation demonstrating the value of quasi-experimental techniques (Glinsmann & Provencher ,2013, Dougherty & Provencher, 2013, Buege et al., 2013). In the case of behavior-based programs, which rely on savings from conservation actions as well as measures, the evaluators' primary objective is to identify a rigorous counterfactual[21] to support estimates of the program effects from consumption data (Todd et al., 2012). The difference in energy use between the counterfactual and the participant group is used to estimate the net savings achieved by the program.

Few program classes have conducted rigorous experimental and quasi-experimental studies to estimate program savings. However, those program classes that have used these methods have

[21] *A rigorous counterfactual is a comparison group that is identical to the treatment group with respect to all energy-related characteristics and behaviors in a given time period. The only difference is that the counterfactual does not receive the treatment. The comparison group should be selected such that customer characteristics and energy-related actions customers take are the same in the pre-treatment time period, and would continue to be the same in the post-treatment period, if not for the program intervention. Theoretically, the ideal counterfactual would be the exact same individual(s) or household(s) in the exact same time period that does not participate in the program (treatment).*

strong evidence of behavioral CIP effects. The most rigorously evaluated behavioral CIP programs utilize experimental and quasi-experimental evaluation designs.

- The behavioral CIP efforts that have the greatest number of *third-party* (independent) evaluations include the following, in order of the number of studies completed: Asynchronous Feedback (Home Energy Reports), Real-time feedback and pricing, Continuous improvement and strategic energy management, Diagnostics, and Benchmarking.

- A number of behavioral CIP classes have reported savings associated with their efforts, however, most have *not undergone third-party evaluations.* As such, these program classes typically report gross savings. These include the following classes, in order of the number of studies completed: Competitions and games, Community-based programs, K-12 schools.

- A select few CIP classes have been evaluated using experimental or quasi-experimental methods to estimate net savings. These include: Asynchronous Feedback, Real-time feedback and pricing, and a select number of community-based programs that have undergone third-party review.

To date, only Residential Asynchronous Feedback programs have evaluated the persistence of savings when treatment is discontinued. Few program classes have measured energy savings over multiple treatment years, and even fewer have measured energy savings after treatment stops. Those that have evaluated persistence without program treatment (e.g., after treatment stops) include the following, most of which are residential programs:

- As we discuss in the next section, about five Home Energy Report programs for which treatment ceased after two years studied persistence in the third year. They found about 80% persistence, on average (Khawaja and Stewart, 2014/2015).

The table below summarizes the approaches typically used by CIP sector and class.

Table 11. Summary of Evaluation Methods Findings by Program Class

Family	Class	Evaluation methods used	Rigor of Evaluations	Persistence evaluated?
Cognition	K-12 Schools	Pre/post billing analysis without a comparison group	Low, and typically gross savings	No
	Continuous Improvement/Strategic Energy Management	Pre/Post analysis without a comparison group, often with engineering adjustments	Medium (Rigorous engineering assessment, but difficult to utilize control group)	No
	Benchmarking	Pre/post billing analysis combined with net-to-gross ratio	Medium	Rarely
Calculus	Diagnostics (On-Site and Remote)	Engineering analysis (deemed savings or building energy software) combined with net-to-gross ratio (survey-based)	Medium (Often net savings, but may rely on deemed savings)	No
	Asynchronous Feedback	Difference-in-differences billing analysis (utilizing Randomized Control Trial)	High (net savings and RCT)	Yes
	Real-time Feedback	Difference-in-differences billing analysis (utilizing matched comparison group)	High (net savings and matched comparison)	Sometimes
Social Interactions	Community-Based	Pre/post analysis or deemed savings/building energy software assumptions, without a comparison group	Low, and typically gross savings	No
	Competition	Pre/post analysis	Low, and typically gross savings	Rarely

8.2 Considering Program Design in Selecting Evaluation Methods

The specific methodology used for evaluation is largely dependent on the program design, namely whether the program utilized an opt-in design or an opt-out design. Opt-in programs are those where customers elect to participate in the program, either through encouragement from program implementers (through marketing or other outreach) or through their own efforts to identify and use tools that support energy savings. Common opt-in behavioral programs are often enabled by technology, such as online interfaces, smart phone applications, and two-way communicating devices. Opt-out programs assign customers to the program, however, customers can "opt out" of the program if they do not want to participate. Also, in opt-out programs, it is generally understood that not all customers assigned to the program ultimately take action in response to the treatment. Common opt-out program models include home energy reports and variable pricing initiatives.

The vast majority of program designs utilize an opt-in approach. When evaluating opt-in programs, the primary goal in identifying a counterfactual is addressing bias associated with self-selection (since customers are volunteering to participate in the program). Evaluators are required to develop a program design capable of addressing the following question: do we feel confident that the effects we are measuring are due to the program and not due to participants' pre-disposition to save energy?

Evaluators should not default to a "gold standard" approach without careful consideration of the complications and confounding factors in implementing the design itself. Rather, evaluators should seek to obtain the highest level of rigor possible while bearing in mind the feasibility of implementing *the evaluation design in field.*

To effectively answer this question, evaluators are tasked with designing a counterfactual (typically in the form of a control or comparison group) that accounts for observable characteristics (such as energy use, demographics, firmographics) and unobservable characteristics (such as pre-dispositions to take action and pro-environmental attitudes) that may drive energy saving practices. Randomized Control Trials (RCTs)[22] are considered the gold standard in counterfactual development. Other purely experimental evaluation techniques (such as Random Encouragement Design (RED) and Recruit and Delay or Deny approaches) can also be considered the gold standard if they are implemented and evaluated properly. The SEE Action Network's protocol "Evaluation, Measurement, and Verification (EM&V) of Residential Behavior-Based Energy Efficiency Programs: Issues and Recommendations" strongly recommends using experimental methods to establish a counterfactual for behavioral programs. By randomizing eligible customers into treatment and control groups, evaluators can feel relatively confident that they have accounted for bias through the randomization process.

However, pure experimental designs are very difficult to implement and maintain in utility program settings for a number of reasons related to the program design. Dougherty and Provencher (2013)

[22] *A Randomized Control Trial (RCT) is an experimental program design in which treatment and control groups are randomly assigned, which results in unbiased program energy savings estimates.*

and Vine et al. (2014) outline a number of challenges in implementing experimental designs in field.

- Experimental designs favor simple treatment approaches. Experimental designs are difficult to implement in practice. For this reason, they favor very simple outreach tactics, such as direct mail, that need to be carefully controlled in the field.
- For various reasons, it is often impossible to randomly assign “subjects” to receive program treatment in the manner of an RCT. The energy efficiency program may be constructed in such a way that subjects self-select into the treatment; or, those who are initially randomly selected may select themselves out of the treatment during the experiment; or they may be assigned to the treatment based on some eligibility criteria that is not random.
- There are extremely important energy efficiency programs that intervene in the market in ways that make it virtually impossible to implement an RCT. For example, one cannot randomly withhold loans from consumers who qualify for them in order to assess the impacts of financing on energy efficiency purchases. One cannot randomly deny selected consumers who are seeking to buy optimizing thermostats at home improvement stores access to them.
- Program design and evaluation are not usually integrated processes, and experiments require experienced evaluators to design and validate the experiment to ensure that the experimental conditions are met and maintained through implementation.

For this reason, Dougherty and Provencher (2013) note that there is no obvious hierarchy to the methods *in practice*; rather, there is an ideal counterfactual approach based on the program design, implementation, and data availability. Thus, evaluators and regulators should not default to a “gold standard” approach without careful consideration of the complications and confounding factors in implementing the design itself. Rather, evaluators should use the most rigorous method possible while bearing in mind the feasibility of implementing the evaluation design in field.

8.3 Overview of Methods

Experimental Design

As noted earlier, experimental design is the only evaluation approach that generates a completely unbiased estimate of program effects. This is chiefly because experimental designs – when evaluated properly – can eliminate selection bias in estimating program impacts. Here, we provide an overview of three experimental design approaches: (1) randomized control trials, (3) random encouragement design (RED), and (3) recruit and delay or recruit and deny.

It is important to note that experimental designs require program models with relatively simple program implementation models and, depending on the anticipated effect size, may require large sample sizes, particularly if the expected energy savings is relatively small (e.g., less than 5%) (Sergici & Faruqui, 2011). Further, recent efforts to implement experimental designs for behavioral programs with technologies (such as home area networks (HANs) or smart thermostats) are difficult to execute in field. A recent HAN evaluation for the Pacific Gas and Electric (PG&E) Company (Nexant, 2014b) found that as many as 18% of their HAN pilot volunteers assigned to the treatment did not successfully connect their device, which could introduce selection bias into the study if the program is not evaluated according to the experimental design. This issue also arose in a Detroit Edison Company's (DTE) dynamic pricing pilot, where the program team was implementing a recruit-and-deny or recruit-and-delay design (see below). In this case, the randomization occurred after enrollment, but *before* people ultimately had equipment installed and participated. The groups ended up being unbalanced in the end, violating the experimental design (DTE 2014a).[23] For this reason, we suggest carefully assessing the feasibility of using experimental designs before implementing.

If feasible to implement, we recommend the following experimental designs:

- **Randomized Control Trials:** A target population of customers (referred to as the sample frame, such as high-income households) is randomly assigned to a treatment or control group. All treatment customers are "defaulted" to treatment, such as receiving a report or a time-based rate. The treatment is then assigned to this group, and the control group does not receive the treatment. The savings effects of the offer are measured across the entire treatment group. RCTs do not measure the individual savings associated with individual households that take action in response to the treatment, as it is understood that some portion of the customers assigned to treatment may not engage with program materials or take action to save energy. The primary drawback to RCTs is that, like experimental designs in general, they are difficult to manage in field for program design models that have multiple engagement strategies or involve multiple technologies. For this reason, home energy reports have been the most successful program model to implement using an RCT because it is fairly simple to ensure that households assigned to a treatment and control group do/do not receive a report as assigned. For other models, RCTs should be used with caution and under the guidance of a qualified third-party evaluator.

---

[23]*See https://www.smartgrid.gov/sites/default/files/doc/files/DTE-SmartCurrents_FINAL_Report_08152014.pdf for how LBNL advised them to rescue it, for future reference.*

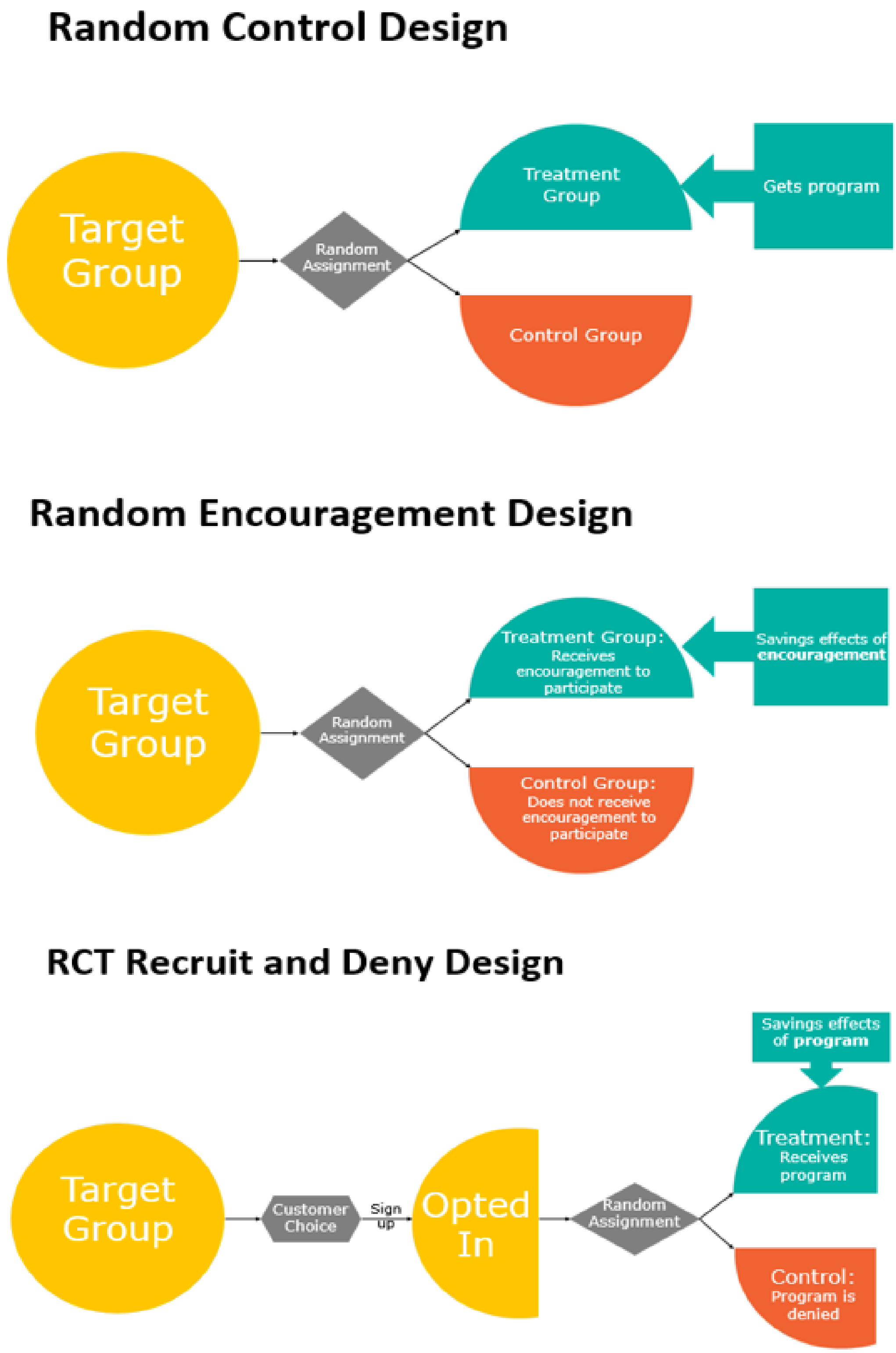
- **Random Encouragement Design (RED):** Members of a target population are assigned to a treatment or control group. The treatment group is *encouraged* to participate in the program (e.g., through direct marketing). The savings effects of encouragement is measured across the entire treatment group (i.e., the effect of marketing). If the participation rate is large enough, one can estimate the savings rate among the households that opt-in, but in practice, participation rates are rarely large enough to support this analysis.

  Random encouragement design (RED) approaches require both large sample sizes and large effect sizes (i.e., large savings relative to energy use) in order to detect savings associated with outreach (i.e., encouragement) as the savings associated with those who act on this encouragement are averaged across all of those who received encouragement (including those who do not act). For this reason, this method is often not implemented successfully in the field and may not be feasible in jurisdictions that have a smaller customer base and cannot meet the sample size requirements. Further, without information on previous participation rates or estimated savings, it can be difficult to determine, in advance, whether or not the RED approach will net statistically significant savings.

- **Recruit and Delay or Deny:** A group of customers who have volunteered and qualified for a program are randomly assigned to a treatment or a control group. The treatment group receives the program. The control group is either denied treatment completely or treatment is delayed long enough to use this group as a control for those who receive treatment immediately. This method measures the savings effects of participation among those who opt-in and receive treatment. This approach is preferred over random encouragement design (in which savings are averaged across everyone who opts to participate) because it does not require extremely large sample sizes to detect program effects through the experiment. Using this method, the treatment effect is measured across only those who receive the program. This is in contrast to the RED approach, which measures the effects of encouragement across the treatment group, only some of whom have take action. If this method is chosen, it is critical to choose and assign groups after customers have gone through as many application and qualification steps as possible, as groups can start to “diverge” after you filter people out for what seems like trivial qualifying details (e.g., type of water heater or type of heating/cooling controls).

  However, this approach may be less favorable from a customer engagement and satisfaction perspective; many utilities and program implementers shy away from this design approach to avoid upsetting interested customers by offering a treatment and then failing to deliver it. For this reason, it is also critical that the communications in the recruitment phase of these projects are careful to manage customer expectations.

Figure 5. Illustrations of Experimental Design Approaches



Quasi-Experimental Design

Quasi-experimental design refers to a method of program evaluation in which a treatment group and a comparison group are defined but households are not randomly assigned to these two groups, resulting in program savings estimates that may be biased. In the case of quasi-experiments, researchers must work to address potential biases in the design of their studies. However, when conducted rigorously, quasi-experimental designs have been proven to produce comparable results to pure experimental designs (in particular, in the case of matched comparison groups, which we discuss later).

Identifying and controlling for selection bias in energy programs may be the single most vexing challenge faced by evaluators. For opt-in programs, self-selection bias is probably the most difficult factor to identify and control. Self-selection bias is introduced when customers are allowed to select or enroll themselves into a participant group, in contrast to opt-out models where all customers are assigned to the treatment group (but can opt out if they desire). This results in a non-probability sample of treatment customers. To account for this bias, a comparison group is drawn to closely replicate the treatment population and the factors ("important differences") that may have led to self-selection, both observable (such as energy use, region, and demographics) and unobservable (such as propensity to participate and internal drive to save energy). The important differences are those that would affect energy usage and energy-use responses to historical/economic events. In an ideal world, we would have sufficient knowledge, time and funding to control for many, if not all, of these factors by applying multiple methods. However, evaluators are most often working under limited information, timeframes and budgets, and must make careful trade-offs to best account for the biases hypothesized to influence participants. For this reason, it is critical to understand the finer details of the program model and theory, how customers were targeted, how it was implemented, and what effects have been observed to date before selecting an approach (Dougherty & Provencher, 2013).

Here, we provide an overview of two quasi-experimental design approaches: (1) matched comparison groups, and (2) variation in adoption method.

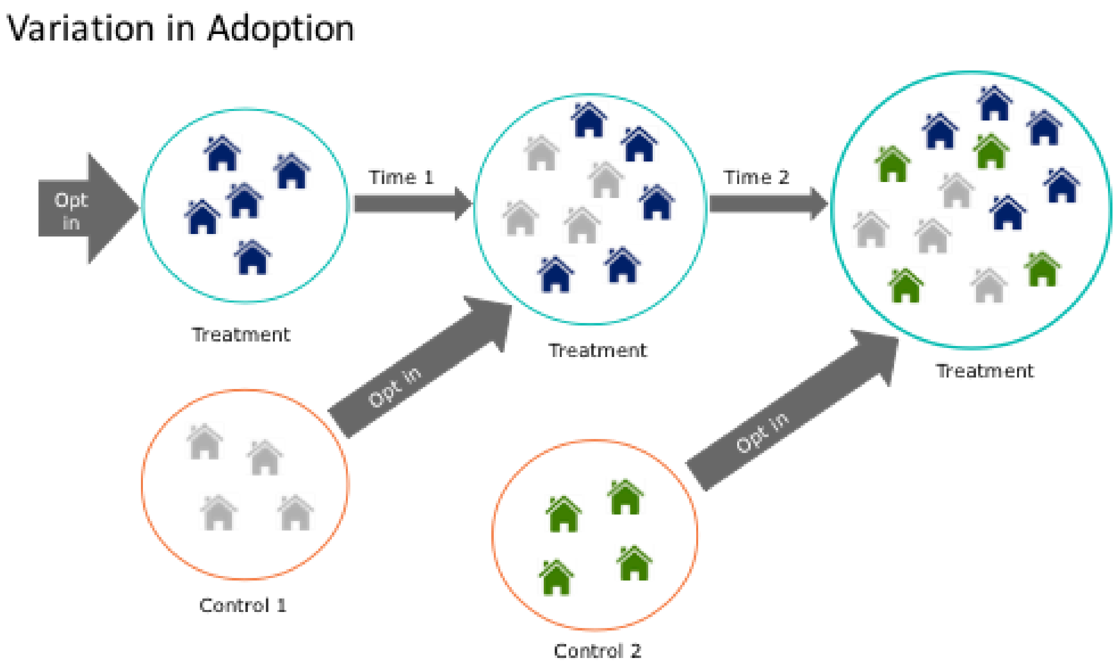
- **Matched Comparison Group:** Participants are matched using observable variables (such as baseline usage, demographics, housing) to non-participants to create a comparison group with similar or exact characteristics. This method measures the savings effect of participation among those who receive treatment. The matched comparison approach in the SEEAction EM&V Guidance (SEEAction, 2012) constructs a comparison group based on matching observable characteristics (energy use for example) under the theory that controlling for these characteristics will control for unobservable characteristics that contribute to selection bias. The matching process can include one or multiple variables used to develop either a one-to-one or one-to-many match of participants to comparison group members, or an overall participant to comparison group balance. This requires pulling matched non-participants on observable variables from a uniform time period prior to program participation.

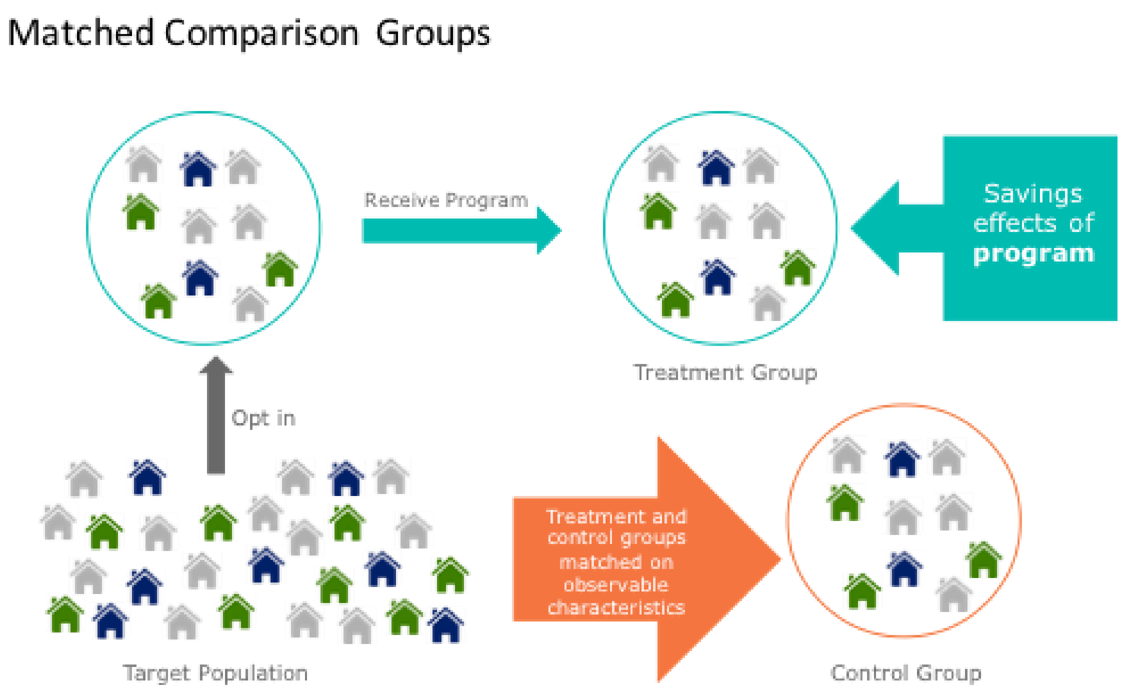
  The matched comparison group method has been frequently used to estimate the impacts of programs in situations where small sample sizes or use of mass media limit the ability to set up an experimental design.[24] Some of the jurisdictions where a matched comparison group method has been used and approved by utility stakeholder or regulatory groups include Massachusetts, Rhode Island (Illume Advising, LLC and Navigant Consulting, 2014), Illinois (Navigant Consulting), California (Itron, 2013) and Minnesota (Illume Advising, LLC, 2014).

- **Variation in Adoption (VIA):** Future participant's pre-treatment usage data are used as the comparison group for current participants on a rolling, month-to-month basis. Additional matching may be used to better align on observable data. The savings effect of participation is among those who opt-in and receive treatment. The approach uses pre-participation billing periods of later program participants ("later adopters") and uses this pre-period data as a rolling comparison group and can be implemented once the program has been implemented. This approach has been used in other evaluation efforts and has been applied in an energy context (Harding & Hsiaw, 2014; Hoynes & Schanzenbach, 2009; Opinion Dynamics 2012). In this work, the later adopters serve as points of comparison until they enter the program as participants, and then are dropped from further analysis in the control group. The later adopters are the comparison group for the earlier adopters. This method has the advantage of minimizing selection bias since ultimately all households opt-in (Todd et al., 2012). However, before using this method, it is critical to assess whether variation in program participation over time may be impacted by external factors such as pronounced weather fluctuations or changes in marketing strategies which can influence the behavior and desire to opt-in among the target group. For example, if participants in month A receive an incentive for signing up and participants in month B do not, then it would be inappropriate to assume that the month B participants can effectively serve as a comparison group for month A participants who likely chose to participate for entirely different reasons. Thus, selection bias is only minimized if the decision of each household to opt in during a particular month is essentially random.

[24] Ho, Daniel E., Kosuke Imai, Gary King, and Elizabeth Stuart. 2007. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. Political Analysis 15(3): 199-236.

Figure 6. Illustrations of Experimental Design Approaches, adapted from Dougherty & Provencher 2013





Preponderance of Evidence Approach

The preponderance of evidence approach typically uses pre-intervention and treatment period ("post") billing data to measure changes in consumption among participants. However, in the absence of a comparison group, the method uses multiple sources of information to produce evidence that the changes detected over time are due, in part, to behavioral CIP efforts. For these efforts, this method should be used as a last resort and may not be considered sufficient evidence for claiming savings in some states. In most cases, this method can be used as a proof of concept, indicating that the design likely impacts energy behaviors. However, without extensive supplementary research, it may be hard to draw a causal link. The California Protocols (Hall, Roth, and Best., 2006; p. 156) note that in "this approach the analyst relies on triangulation from multiple data sources to draw conclusions about the presence and attribution." While the protocols refer to this approach specifically in reference to market effects, it has been used in many states to examine impacts associated with EERS programs.

For program models that use this approach, we recommend that program implementers work closely with evaluators to identify, track, and analyze data throughout the program cycle in order to measure changes that may be related to the program. In addition, the table below illustrates participant net effects that may be valuable in demonstrating causality.

Table 12. Levels of Rigor Outlined to Estimate Participant Net Effects as Presented in the California Protocols (Hall et. al, 2006; p. 36)

Rigor Level	Minimum Allowable Methods for Participant Net Impact Evaluation
Basic	1. Participant self-report
Standard	1. Participant and non-participant analysis of utility consumption data that addresses the issue of self-selection. 2. Enhanced self-report method using other data sources relevant to the decision to install/adopt. These could include, for example, record/business policy and paper review, examination of other similar decisions, interviews with multiple actors and end-user, interviews with mid-stream and upstream market actors, Title 24 review of typically built buildings by builders and/or stocking practices. 3. Econometric or discrete choice with participant and non-participant comparison addressing the issue of self-selection.
Enhanced	1. "Triangulation" using more than one of the methods in the Standard Rigor Level. This must include analysis and justification for the method for deriving the triangulation estimate from the estimates obtained.

8.4 Methods Trade-offs

As discussed earlier, the specific methodologies used for a given program depends largely on the program design. In the table below, we provide a "cheat sheet" for determining which methodologies may be the most suited to a given behavioral program based on the design elements and assuming that expected savings per premise is low. However, it is important to note that this is an oversimplification, and the specific design should be determined in consult with a professional evaluator.

Table 13. Summary of Evaluation Methods Used by Program Class

Best for programs that...	RCT	RED	Recruit-and-Delay or Deny	Matched Comparison	Variation in Adoption	Preponderance of Evidence
Use an Opt-Out Model	✓	×	×	×	×	×
Use an Opt-In Model	×	✓	✓	✓	✓	✓
Large percentage of customers expected to participate (opt-in, or not opt-out)	✓	✓	=	×	✓	=
Small number or percentage of customers expected to participate	×	×	✓	✓	×	✓
Equipment requires installation or is complex	×	×	×	✓	✓	✓
Wide range of recruitment tactics or sign-up options	×	✓	✓	×	×	✓

Key: ✓ = Advisable for this program design element

× = Not advisable for this program design element

= = Possible, depending on program design

8.5 Cross Program Participation

In our review, few program evaluations investigated how many participants in the behavioral program participated in other utility rebate programs ("cross program participation"), and whether their participation rate and savings through other utility rebate programs was over and above non-participants' savings. Only Home Energy Report evaluations systematically analyze program cross-participation, for the purpose of deducting "double-counted savings" from the behavioral program claimed savings, since incremental savings achieved through other utility rebate programs are typically claimed by those rebate programs. The table below summarizes incremental savings achieved through other utility rebate programs, defined by the equation below:

Program design may promote cross-program participation. In the last column of the table below, we indicate what each program class's potential for driving cross-program participation may be, based on the program design.

Table 14. Behavioral Program Metrics and Potential to Drive Cross-Program Participation

Family	Class	Incremental savings typically measured?	Percent of behavioral program net savings that are incremental energy efficiency (EE) program savings		Potential to drive cross-program participation
			Electric	Gas	
Cognition	K-12 Schools	No	NR	NR	Low
	Continuous Improvement/ Strategic Energy Management	Sometimes (some evaluations remove EE program savings to estimate O&M savings only)	NR	NR	High (by design)
	Benchmarking	No	NR	NR	Low
Calculus	Diagnostics (On-Site and Remote)	Rarely			High (by design – audit/assessment programs are typically designed to channel customers to other programs)
	Asynchronous Feedback	Yes	Average 5.6% Range -2% - 24%	Average 3.7% Range -8% - 10%	Medium
	Real-time Feedback	No	NR	NR	Low
Social Interactions	Community-Based	No	NR	NR	High (by design – audit/assessment programs are typically designed to channel customers to other programs)
	Competition	No	10% (Duke Smart Energy Now)	NR	Low

[a] Percent of behavioral program net savings that are incremental EE program savings.; NR=Not Recorded.

For the purposes of determining the effectiveness of a given behavioral program, we recommend that the programs estimate the impact of program channeling or cross-program participation. This will serve as another metric upon which to estimate the programs' effectiveness.

Section 3: Energy Efficiency Behavioral Programs: Measure Life and Average Savings Methods Discussion

9.0 Behavior CIP Measure Life & Cost of Saved Energy

Administrators and regulators in many states are actively considering the measure life associated with behavioral programs. Unlike equipment-based programs, where a measure is installed in one program year and generates an estimated savings value across multiple years, behavioral program savings are generated from a wide range of actions across multiple end uses, each with varying impacts over time. Behavioral programs generate savings from both installed equipment and "conservation behaviors" which rely on the choice of individuals within households and organizations to change the way they use equipment. Conservation behaviors, in particular, are at risk of decaying over time, as customers may stop taking actions without continued prompting by programs. With this knowledge, program administrators made a conservative assumption about the measure life of behavioral programs when these programs first launched, adopting a one-year measure life. This one-year measure life assumes that customers generate savings only when the program is active.

Since their initial launch as resource programs, longitudinal research and persistence studies on Residential Home Energy Report (HER) have shown that a one-year measure life may be too conservative. Recent studies (describe more in this section) indicate that savings do persist for HERs without continued treatment (albeit with decay) among participants that received a HER for at least two years before it was discontinued.

In this section, we discuss the measure life findings from various studies, and their implications, for the Minnesota DER in planning for, and claiming, behavioral program savings against its CIP savings goals.

9.1 Context and Definitions: Integrated Resource Planning and Cost-Effectiveness Testing

It is important to consider the context in which program administrators (PAs) claim energy savings in Minnesota (and other jurisdictions) in order to understand challenges associated with extending the measure life of HERs. The three main venues in Minnesota include:

- Integrated resource planning (which sets energy and demand savings goals);
- Conservation Improvement Program, and
- Shared Savings DSM financial incentive mechanism.

Integrated Resource Planning and Achievement of CIP Goals

Each program submits a claim for savings associated with the measures that were installed in each calendar year that counts towards establishing a savings goal for each program.

- The savings claimed toward CIP goals in Minnesota are "first-year savings" only – i.e., the savings obtained in one year of the program only. For equipment or "asset-based" programs, these savings are measured as average annual energy savings of the installed measures. Utilities receive annual energy savings for each piece of installed equipment, but only for the calendar year in which equipment is installed. No lifetime savings are credited. Thus, measure life is not part of the calculation of first-year savings for asset-based programs.

- Note that though these savings are referred to as “first-year savings”, they are claimed for each year of multi-year and long-standing programs. The “first-year” designation means the savings generated over the first year after a measure was implemented. First-year savings are credited to the calendar year of implementation. For multi-year behavioral programs, for example, “first-year savings” can be claimed for each year of program implementation under the one-year measure life assumption. Many states including Minnesota credit energy efficiency (EE) programs with only first-year savings, although lifetime savings are factored into the standard benefit-cost tests for all other energy efficiency programs.

- For Minnesota’s integrated resource planning, the optimal amount of DSM resources for a utility’s expansion plan is evaluated by comparing the costs (present value of social costs, equal to the present value of revenue requirements plus environmental costs) of thousands of different expansion plans with varying amounts of both supply-side and demand-side resources. Thus, the amount of DSM resources that are found to be cost-effective is dependent upon the lifetime cost of energy savings *and* how the DSM resources interact with all of the existing and potential new utility resources.

Utility Cost-Effectiveness Tests

In Minnesota, investor-owned utilities are required to calculate more comprehensive **cost-effectiveness metrics** than “program cost of saved energy”. Administrative rules for the “determination of reasonable investment” require that CIP program cost-effectiveness be analyzed "from the utility, ratepayer, participant, and societal perspectives".[25] In Minnesota, this translates into the Program Administrator Cost Test (PACT, also known as the Revenue Requirements or “Utility Cost Test” (UCT)), the Ratepayer Impact Measure (RIM), the Participant Cost Test (PCT), and the Societal Cost Test (SCT).

- The lifetime savings achieved by each year of program investment are accounted for in all four cost-effectiveness tests. When measuring “lifetime savings,” the measure life of a given piece of equipment is used to determine how long that measure will generates savings throughout its “effective useful life.”

- However, asset-based programs and behavioral programs have different methods for estimating lifetime savings that affect the assessment of lifetime savings and thus “reasonable investments.” For asset-based programs, the measure life is often obtained through market research or engineering studies, and reflects the typical life cycle of a piece of equipment before it needs replacement. For behavioral programs, the measure life is not estimable in this manner, and assumptions have largely been based on theories of what actions drive savings through behavioral programs. As stated above, in the early years of behavioral programs when little information was available on what actions customers were taking, PAs conservatively assumed a measure life of 1 year.

We refer to the group of utility tests uses to determine “reasonable investment” as “cost-benefit tests” and the resulting ratios as “cost-benefit ratios”. The cost of saved energy, which can be calculated from first-year savings and program costs for a given program year, can be considered a

[25] *Minnesota Administrative Rules 7690.1200, Subp. 1(C) (https://www.revisor.mn.gov/rules/?id=7690.1200)*

proxy for cost-effectiveness, though it is not a cost-effectiveness metric in the strict sense where administrative rules require that tests consider the full range of benefits and costs. In this document, we refer to “cost of saved energy” by its name, and when we mention “cost-benefit tests,” we are referring to the more comprehensive set of utility costs tests. The term “cost-effectiveness” is used more generally, to refer to the family of possible metrics such as cost of saved energy and cost-benefit tests that can be used to understand a program’s performance.

Considering Behavioral Programs in Integrated Resource Planning

An important goal of the State of Minnesota’s Conservation Improvement Plan (CIP) is to generate reliable long-term efficiency savings to be used in resource planning. The State has adopted the “Average Savings Methodology” (ASM) to measure behavioral CIP savings as part of its energy efficiency as a resource (EER) programs.

This method was developed and adopted to address the following concerns: (1) avoid disproportionately weighting the value of behavioral CIPs relative to asset-based programs, which would undermine the lifetime savings generated through CIP, decreasing total emissions reductions and the value of CIP as a resource in integrated resource planning and (2) inflate the DSM financial incentives claimed by investor-owned utilities. In this section, we discuss the ASM approach and compare it to other approaches used to account for behavioral CIP efforts with a multi-year measure life assumption.

Notably, the ASM only modifies how first-year savings are apportioned for the purpose of claiming savings. Net benefits, used in cost-effectiveness tests and investor-owned utility DSM financial incentive calculations, can still be calculated using 100% of observed savings for each program year. In the following sections, we discuss different approaches to claiming savings and discuss how measure life assumptions affect cost of saved energy metrics.

9.2 CIP Measure Life Consideration

Traditionally, CIP behavioral savings are measured and claimed annually using a one-year measure life. Under the one-year measure life assumption, the PA can claim all savings observed in a given year, per the discussion above. HER CIP programs, in particular, use a cohort approach where groups of customers enter the program at the same time and continue to receive reports at the same rate over time. For each year that the program is implemented, it receives the full credit for the measured savings regardless of whether some portion of the savings are due to actions taken in previous program years. However, over the course of multiple years of treatment, a portion of the savings accrued in year one persists into future years.

Recent research has demonstrated that HERs show a relatively consistent savings decay rate once the program has ended. As shown in Table 14, decay rates range from 11% to 32% annually as cited by Khawaja and Stewart (2015). Across all studies, these data suggest that a one-year measure life is not accurate for this specific program type.[26]

---

[26] It is important to note here that no such studies have been conducted for other program models within this class or across CIP efforts documented in this paper.

Table 15. Post-treatment Savings Estimates for HER Programs

Authors	Utility or Service Area	Frequency of Reports	Number of Treatment Months	Number of Post-Treatment Savings Analysis Months	Key Findings About Savings Decay
Allcott and Rogers (2014)	Upper Midwest	Monthly and Quarterly	24–25	26	Average annual savings decay of 21%
	West Coast	Monthly and Quarterly	24	29	Average annual savings decay of 18%
	West Coast	Monthly and quarterly	25–28	34	Average annual savings decay of 15%
Integral Analytics (2012)	SMUD	Monthly and Quarterly	27	12	Savings decay of 32% one year after treatment stopped
DNV-GL (2014a)	Puget Sound Energy	Monthly and Quarterly	24	36	Average annual savings decay of 11%
ODC (2014)	National Grid - Massachusetts	Bi-monthly and Quarterly	12-24	10	Reduced treatment leads to reduced observed savings. Effect is sharper for gas cohorts.

Source for all but ODC results: Table 1 from Khawaja, M.S. and J. Stewart. *Long-Run Savings Cost-Effectiveness of Home Energy Report Programs.* The Cadmus Group. Winter 2014/2015.

Source for ODC results: Arnold, H. Massachusetts Cross-Cutting Evaluation: Home Energy Report Savings Decay Analysis. Opinion Dynamics. September 2014.

To date, four states have extended the HER measure life beyond one year: Connecticut, Maryland, New Hampshire, and Washington. Notably, the same measure life is not used by these states (with three years being the most common). Drawing on this research, Khawaja and Stewart (2015) argue for calculating an extended measure life based on the average annual savings decay rate. Rather than recommending a specific value, they propose a formula for calculating effective useful life based on program parameters and an assumed decay rate.[27]

Effective Useful Life = Lifetime Savings/First Year Savings where:

Lifetime Savings = First Year Savings/ $(\delta+\alpha - \delta*\alpha)$ where:

δ is the annual decay rate and α is the annual attrition rate.

[27] The authors provide a formula that calculates effective useful life as a function of first-years savings, an annual decay rate that is assumed to be the same for all years of the program, and average customer attrition (i.e., customers moving).

Applying the Khawaja and Stewart (2014/2015) approach using an annual decay rate of 20% and customer attrition rate of 7%, results in an effective measure life of 3.9 years, much higher than the one-year measure life assumption. Notably, a HER measure life greater than one year changes our theory of how savings from the programs accrue, and in turn, how savings in program years two and following should be attributed to investment in each of those years. The one-year measure life approach leads to possible misattribution of savings against the costs for operating the program after the first year, possibly distorting the cost of saved energy estimates for CIP efforts relative to asset-based programs with first year upfront costs. For this reason, recent evidence of behavioral program persistence (and thus a measure life greater than one year) has prompted states to reconsider these assumptions.

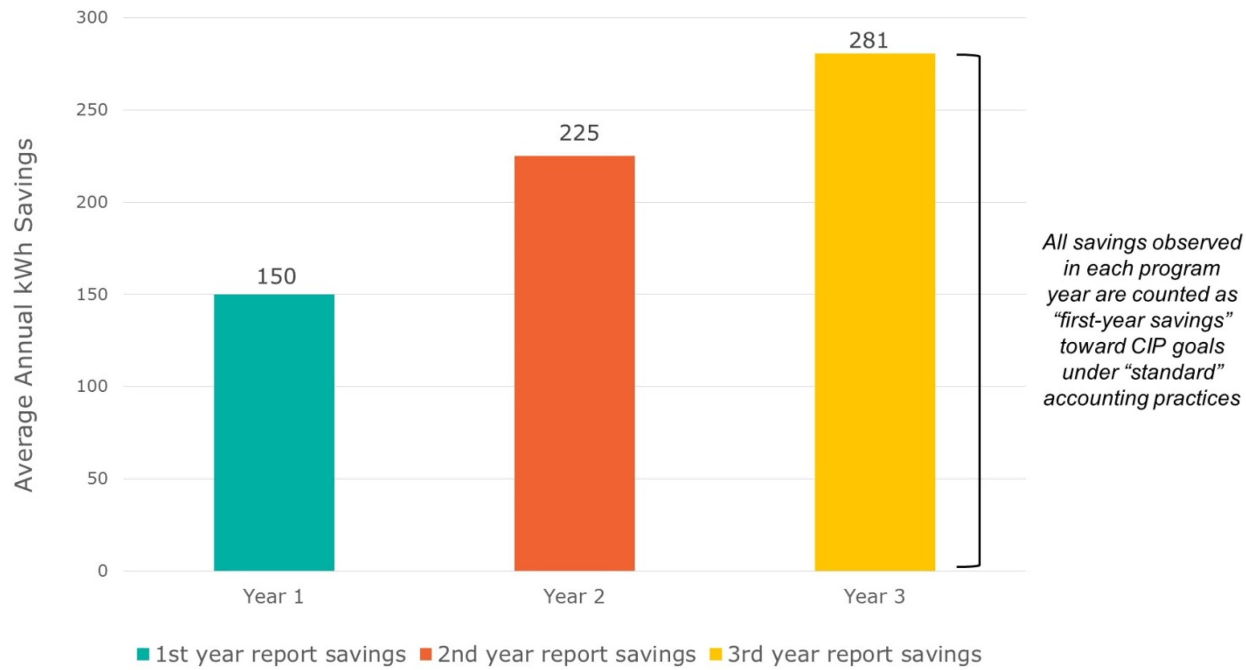
9.3 The Standard Approach to Claiming Savings

In this section we discuss the standard approach to measuring and claiming CIP program savings, and the distortions this approach may create for cost of saved energy and cost-effectiveness comparisons. This approach is used in most jurisdictions, and was used in Minnesota prior to the ASM. In subsequent sections we discuss alternative approaches designed to mitigate some of the concerns with this approach.

In standard behavioral CIP programs, only the first year savings are counted toward CIP goals. More specifically, savings are calculated in each program year, and all savings achieved by participants in a given program year are associated with program efforts and investments in that year.

The figure below provides an illustrative example of the standard method of accounting for savings in a three-year HER program, where a single group of customers (or a cohort) is treated with a report. For each year that this group of customers receives a report, the program savings are measured through a billing analysis. Savings have been shown to increase each year for the first three years of the program, after which they tend to level off (not depicted here).

Figure 7. Example of Average Annual Household Savings Measured for a Residential Home Energy Report Program Cohort over Three Years of Treatment



This approach accurately accounts for all "first-year savings" achieved by the program – i.e., all first-year savings are claimed at some point. However, under the standard assumption of a one-year measure life, this approach assumes that all savings measured in each program year (1-3) are newly acquired, first-year savings. As a result, this approach also produces distorted cost of saved energy results if savings persist from year to year. Residential HER persistence studies (presented in Table 15) have demonstrated that this particular form of CIP program savings persist once participants stop receiving reports (albeit with decay).

For example, if some of the savings from year one persist into year two, attributing all of the savings achieved in Year 2 to Year 2 expenditures would inflate the effect of those Year 2 expenditures (because some of those savings may be "carry-over" from savings the program encouraged in Year 1 by the same group of participants). Thus, savings that are claimed against program costs each year are misattributed to the costs spent in that year. This differs in comparison to other programs, where we would expect new participants over time, such that all observed "first year savings" in each year are independent from savings gained in prior years.

Thinking about the implications for our example above, this means that the savings gained in year two include savings acquired in year one, and savings gained in year three include savings gained in years one and two. For these reasons, some state entities have attempted to account for this issue using various methods to adjust savings across the measure life. Methods that are currently used to attribute savings across program years include:

1) **Average Savings Methods:** Mandated in Minnesota (and adopted by Xcel in Colorado), the average savings method requires that utilities claim one-third of the observed savings each year in a three year planning period. After three years, the utility has claimed the annual average observed savings for the three-year period. This method effectively sets a three-year

measure life for behavioral programs, with the utility claiming one-third of the total observed savings over the three years

2) **Avoided Decay:** Used by Connecticut Light and Power (CL&P), the Avoided Decay approach assumes an annual reduction (or decay) in savings across the measure life for savings acquired in the first year, while accounting for savings gained in each subsequent year through "new" incremental savings gains and avoided losses (or avoided decay) due to continued treatment by the program. This method aligns with that proposed by Khawaja & Stewart (2014/2015); we provide a more detailed example below.

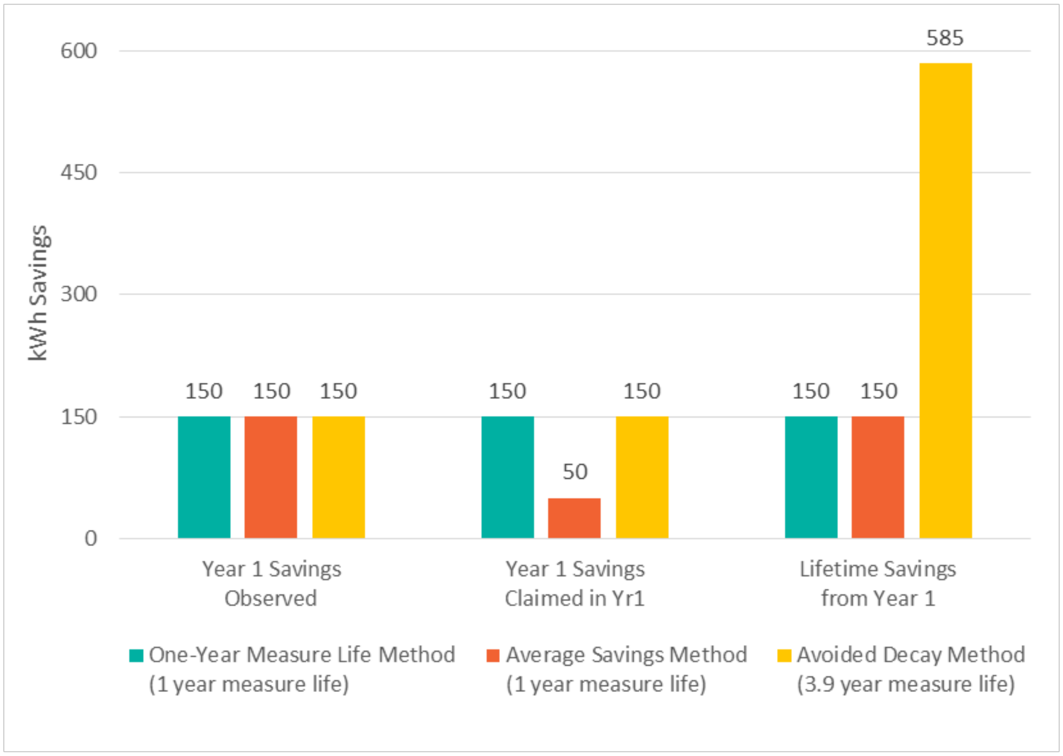
We describe and compare both the ASM approach and the Avoided Decay approach in this chapter. In addition to these two approaches, there is also the Incremental Savings approach:

3) **Incremental Savings:** Utilized in the Northwest by Puget Sound Energy (PSE) and under consideration in the Regional Technical Forum (RTF), the incremental savings approach (our term) allows for program saving to persist for two years after the first year of treatment, while also allowing implementers to claim savings associated with incremental gains in savings over previous years. *Notably, this approach is unique in its application to PSE's planning process and for this reason we do not describe it in detail in here.*

4) **One-Year Measure Life:** Finally, we refer to the standard, status quo method described in the previous section as the "One-year Measure Life" method. This method assumes a measure life of one year and claims all savings observed in a given year toward that year's efforts. As noted above, this method does not address the problem of attributing savings achieved in years two and later to prior year efforts.

Note that while the Average Savings Method claims savings for year one over three years, it is merely partitioning the first year savings over three years, not persisting savings over an extended lifetime. Applying an extended measure life to year one program savings produces mush higher lifetime savings. Figure 8 shows observed, claimed, and lifetime savings under all three approaches.

Figure 8. Comparison of Annual and Lifetime Savings for the First Year of a Residential Home Energy Report Program Cohort



9.4 The Average Savings Method Approach

The ASM was developed in a stakeholder process in 2010 to address the concern that behavioral programs could undermine the total lifetime savings generated through CIP. Further, the ASM addresses the concern that behavioral programs could inflate investor-owned utility DSM financial incentives, as described below. The ASM method proposes that:

1. Each behavioral program participant receives feedback for three years minimum, and savings are measured every year,
2. Measure life for the purpose of claiming savings toward CIP goals is three years. First-year savings claimed for any given program year are equivalent to observed in-year savings divided by three, so that after three years, the utility has claimed the annual average measured savings. This is conceptually similar to asset-based programs under Minnesota's EERS, where a utility claims only the first-year savings over the lifetime of an asset measure. The first-year savings for an asset measure are equivalent to the annual average energy savings.
3. The lifetime savings of the behavioral program are used for calculating net benefits in cost-benefit testing consistent with non-behavioral programs. For the purpose of cost-benefit testing, the measure life is still assumed to be one year, so there is essentially no change to how savings are considered in cost-benefit tests relative to asset programs.

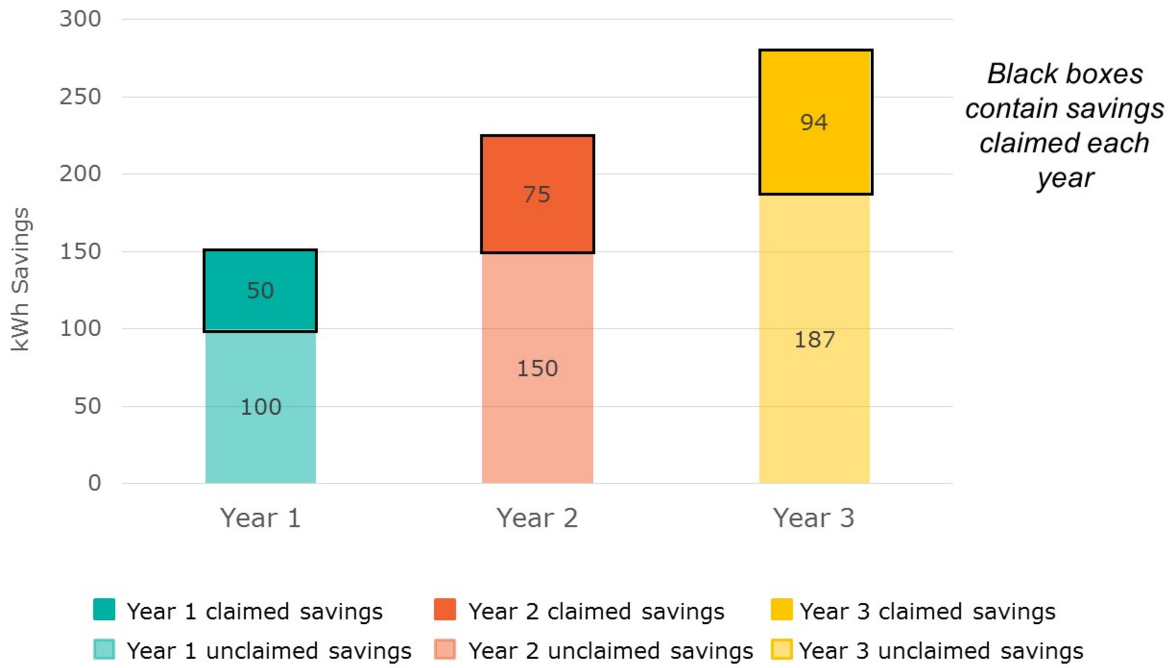
The focus of our discussion is point #2: The ASM addresses the issues caused by the perceived "double-counting" of first year savings by the same participants each year of the program by dividing each first-year savings value[28] by three and applying this number to each of the three years of the treatment in their triennial CIP plans. For each year that the program reports savings in its annual status report, the utility is credited with the actual measured savings for that year divided by the lifetime of the project. In Minnesota, the planned lifetime of the behavioral project is three years. If we use the first-year savings shown in Table 15 (150 kWh) and assume a measure life of three years, the ASM approach would require that the program claim one-third of the measured savings in each of the three years of program implementation. The figure and table below illustrate this method. For more detail, please see Minnesota DER decisions "In the Matter of Inclusion of Behavioral Project Savings in Energy Conservation Improvement Programs and Shared Savings Demand-Side Management Financial Incentive Calculations", dated February 1, 2012 and April 26, 2012.

Table 16. Example of How Savings Measured for Each Year in Figure 9 are Claimed under the Average Savings Method Over Time

Average Savings Method	Year 1	Year 2	Year 3	Year 4	Year 5
Observed Savings	150	225	281	225 *(=80% of previous year)*	180 *(=80% of previous year)*
1st year report savings	50	50	50		
2nd year report savings		75	75	75	
3rd year report savings			94	94	94
Savings Claimed toward CIP Goals	50 *(=1/3 of observed)*	75 *(=1/3 of observed)*	94 *(=1/3 of observed)*	0	0

28 First-year savings are typically calculated for each cohort's first, second and third program year, through difference-in-differences billing analysis. Though the savings might apply to the second or third year of the program, we refer to these observed savings as "first-year savings" due to how they are considered in accounting.

Figure 9. Example of How Savings Measured for Each Year in Table 1 are Claimed under the Average Savings Method



Savings that are measured, but not claimed toward first-year savings goals (e.g., 2/3 of savings observed in each year), are considered in cost-benefit ratio calculations, but do not count toward 1.5% savings goals in any program year.

9.5 The Avoided Decay Approach

In this section, we discuss a relatively new approach to attributing savings over time that incorporates the latest information on the persistence of savings beyond the program year. The Cadmus Group recently published a study entitled *"Long-Run savings and Cost-Effectiveness of Home Energy Report Programs"*[29] that makes an argument for adjusting the savings for CIP efforts over time to account for "persistence" stemming from two sources: (1) persistent savings over time as a result of installed equipment or habituated behaviors, and (2) avoided savings decay, which accounts for the retention of behavioral savings due to continued treatment.

The figure and table below illustrate this method.

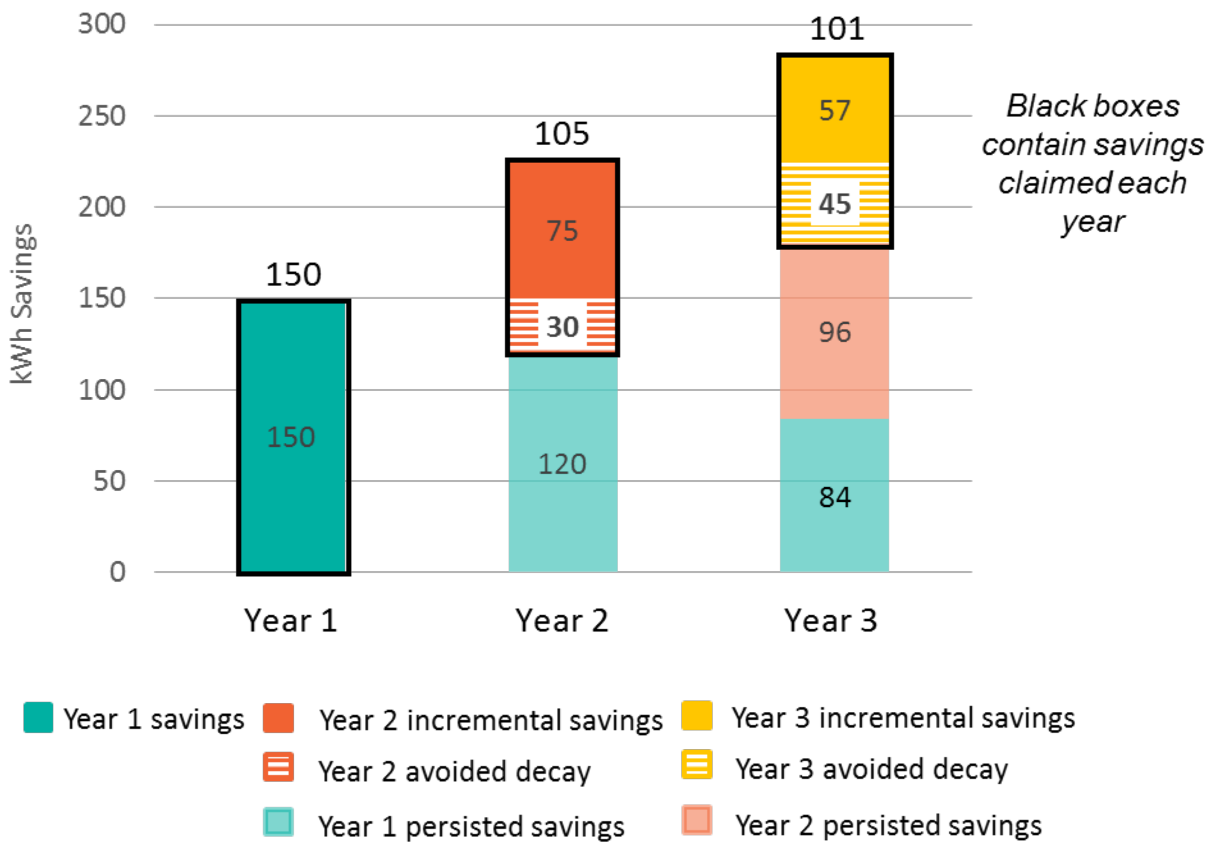
1. First, some savings are achieved, and then persist, from one year to the next. These savings are "persisted savings" in the figure below, and can be estimated with a decay rate (e.g., last column of Table 15, above).

2. These persistent savings are not claimed toward first-year goals in the years after they are achieved. However, they may be counted in cost-benefit tests.

[29] *The Cadmus Group.* Long-Run Savings and Cost-Effectiveness of Home Energy Report Programs *Prepared by M. Sami Khawaja Ph.D. and James Stewart Ph.D. Winter 2014/2015*

3. Second, the method accounts for savings that are maintained ("avoided decay") due to continued treatment, as well as the "incremental savings" gains achieved in a given year. Savings that the program either maintains or adds in each of years two and three are attributed to treatment in years two and three, and are the basis of savings claims for those years. For more detail, see Khawaja and Stewart (2014/2015).

Figure 10. Example of Sources of Savings in Each Program Year[30]



The table below summarizes how savings are calculated and claimed in each year. In this example, we assume 80% persistence of savings from year to year. Claimed savings from years two and three are the sum of what program efforts either maintain ("avoided decay") or add ("incremental savings") in each year. In practice, these numbers are not calculated individually. Instead, evaluators use the persistence assumption (here, 80%) to estimate what prior program years contributed to year 2 and 3 savings, and take the remainder as what the program can claim against CIP goals in each of years 2 and 3. The table below details the inputs to these calculations.

[30] *Image Adapted from the following source: Cadmus. Long Run Savings and Cost-Effectiveness of Home Energy Report Programs. Prepared by M. Sami Khawaja, Ph.D. and James Steward, Ph.D. Winter 2014/2015.*

Table 17. Example of How Savings Measured for Each Year in Figure 1 are Applied under the Avoided Decay Method over Time

Avoided Decay Method	Year 1	Year 2	Year 3	Year 4	Year 5
Observed Savings	150	225	281	225 *(=80% of previous year)*	180 *(=80% of previous year)*
Savings due to year 1 efforts	150	120 *(=80% of previous year)*	96 *(=80% of previous year)*		
Savings due to year 2 efforts		105 *(remainder)*	84 *(=80% of previous year)*	67 *(=80% of previous year)*	
Savings due to year 3 efforts			101 *(remainder)*	81 *(=80% of previous year)*	65 *(=80% of previous year)*
Savings Claimed toward CIP Goals	150	105	101	0	0

9.6 Comparative Cost-of-Saved-Energy Under the Standard, ASM, and Avoided Decay Approaches

Each of the three outlined approaches outlined have a varying impact on the cost effectiveness of the program. We summarize the implications by (a) using the cost of saved energy ($/kWh) as a proxy for cost-effectiveness (because it is a common metric for comparing programs in integrated resource planning), and (b) assuming a cost of $12 per participant per year. When we compare the cost of saved energy for each measure life adjustment approach, we see that:

- **The One-year Measure Life (Standard Approach)** draws down the cost of saved energy in years two and three. In our opinion, the cost of saved energy in years 2 and 3 is artificially low, because some portions of those savings were due to previous program year efforts.

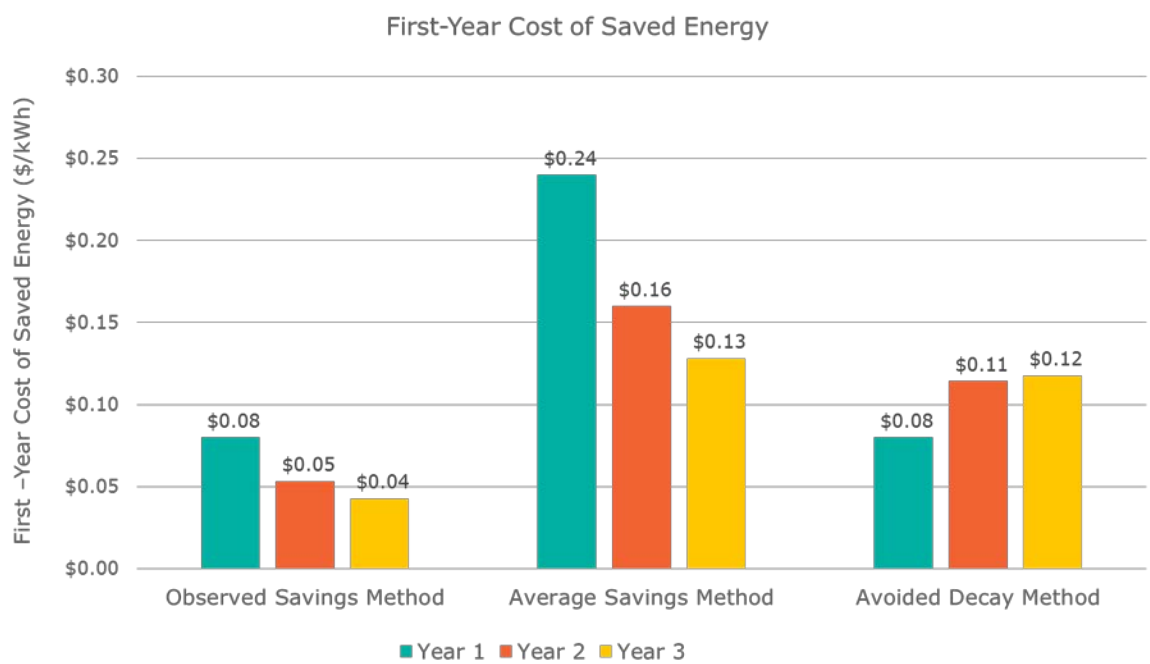
- **The ASM** method significantly increases the first year cost of saved energy with steady improvement in subsequent years. In our opinion, the first-year cost of saved energy is artificially high, because the program actually did realize all of the observed savings in that first program year.[31] The results also show improving cost of saved energy over time, whereas, in the face of savings persistence, the cost of saved energy likely worsens over time

[31] In this particular example the cost of saved energy also appears high in years two and three relative to the cost of saved energy from the avoided decay approach, but we do not wish the reader to conclude from this example that the cost of saved energy would be inflated in years two and three for other programs or sets of assumptions.

(as more and more savings persist from year to year, but program expenditures remain constant).

- **The Avoided Decay** approach demonstrates a reasonable first year estimate based on the measure savings, with a gradual increase in the cost of saved energy over time. In our opinion, these estimates are a more accurate reflection of what savings were achieved by each program year's expenditures. Given that the portion of savings that are either "avoided decay" or "incremental savings" due to each subsequent year's efforts grow smaller and smaller as previous year savings persist, it makes sense that the first-year cost of saved energy grows over time. The avoided decay approach is the only approach that reflects this trend.

Figure 11. Comparative Cost of Saved Energy by Measure Life Accounting Methods



9.7 Accounting for CIP Efforts with Rolling Enrollment

It is important to note here that these methods would require adjusted accounting methods for CIP efforts with rolling enrollment when compared to the HER cohort model. Rolling enrollment is characteristic of most opt-in, online, and real-time feedback programs. As customers trickle in over time, the age of the savings "acquired" can vary significantly for each year the program is implemented. For example, if 100 participants enter over the course of year one, and another 100 over the course of year two, the new participant savings acquired in year two, under these methods, would be inappropriately treated as second year savings when, in fact, 50% of the savings measured in year two are essentially first-year savings. We recommend addressing this by breaking participating customers into cohorts every six to twelve months to more accurately account for first year vs. subsequent year savings. This will more accurately reflect the "age" of the savings acquired in order to attribute savings to each year's efforts (i.e., program expenditures).

It is important to note that there is not sufficient research in place to support the use of the Avoided Decay approach for opt-in and rolling enrollment models. To date, none of these programs have developed a persistence analysis to examine how savings decay over time. While the discount methods used in the Avoided Decay approach may be better than an arbitrary assignment, future research should be conducted to identify a more appropriate decay rate for opt-in program models.

9.8 Discussion and Recommendations

The benefit of the ASM is that it reduces "double-counting" of persisted savings that would otherwise occur under the standard one-year measure life approach. It is also relatively simple and standardized across all CIP programs and program administrators throughout the state. In this way, it is predictable for program administrators. However, in reviewing the ASM against other models, our assessment is that the ASM results in a) undercounting the savings achieved in the first year, b) difficulty in accurately attributing savings to year two and three efforts, and c) an inaccurate assessment of the cost of saved energy for the program over time. In particular, the ASM may under-estimate savings in early years and distort the relationship between the cost of saved energy and time (making programs seem more cost-effective over time. In reality, if there are persistent savings, the cost of saved energy would worsen over time if investments remain the same).

With new insight into CIP persistence since the ASM decision, we recommend that the State of Minnesota consider refining the ASM approach for HER programs, based on actual data on how savings decay over time. Specifically, we recommend considering the Avoided Decay approach, or its underlying logic, to develop a revised ASM for claiming first-year savings.

In addition to more accurately reflecting first-year savings attributable to each year's spending, the Avoided Decay approach provides a solid framework for making program design decisions. Under the Avoided Decay theory, at some point the avoided decay and incremental savings will no longer produce savings great enough to justify the investment, signaling the end of life for the intervention with that group of customers. This is important because if policies determining how long to continue investing in behavioral CIP efforts are driven by cost-effectiveness results, the models must be run so that it accurately reflects in-year savings to costs. Further, this approach may also be useful in developing deemed savings for behavioral CIPs by year of treatment. The state of Michigan has deemed first year savings for behavioral programs but no one has looked at deeming savings over the course of multiple years of treatment.

While this approach is relatively straightforward for opt-out HER program models where all customers are brought in at the same time, it is more complex for opt-in models where customers may join the program at any point over the course of a treatment year.

Finally, we caveat that the assumption of 80% persistence that we used for examples in this paper is based on only a handful of studies in other jurisdictions. Most of these studies stopped treatment after about two years of delivering HERs, and observed what happened in the third year. Applying this number as the de facto persistence rate for all program years assume that (a) these studies are applicable to Minnesota programs, and (b) persistence is consistent across program years. We recommend that Minnesota continue to monitor research in this area, and consider funding experiments to measure persistence in HER and opt-in program models.

APPENDIX I: DETAILED METHODOLOGY

Approach to Defining Behavioral Programs

To commence our study, we began by developing a definition of behavioral programs. For the purposes of this study, programs met our definition of "behavioral" if they specifically used behavioral strategies to influence behavior change related to energy consumption, and if they were evaluable for energy savings. The full definition of what constitutes a behavioral program for inclusion in this study is described in greater detail in Section 4, below. This definition was submitted to the Minnesota DER on January 27, 2015 in a memorandum. Upon agreement, this definition functioned as inclusion criteria for programs we reviewed during the literature review. Programs that did not meet this definition were excluded from this literature review.

Approach to Defining the Taxonomy

In addition to defining what constitutes a behavioral program, we also developed two taxonomies specific to the residential and commercial sectors. Each taxonomy contained four levels that differentiate types of behavior-based efforts. It is important to again note that programs were only classified in our taxonomy if they also had demonstrated savings-related metrics. The taxonomy is structured to enable comparison of behavioral efforts across program classes – such as community-based efforts or "asynchronous" feedback[32] like Home Energy Reports - rather than behavioral intervention strategies or tactics. This is an important element of the taxonomy: we chose this orientation that categorizes efforts by program class because this is how efforts are typically (or could be) delivered by a utility to its customers, rather than by more specific components or strategies of these programs (for which it is typically difficult to isolate savings). We also present benchmarking results at the program class level to enable stakeholders reviewing this study to make portfolio allocation and investment decisions at the program level. However, as we discuss further below, in some cases it was necessary to (a) further differentiate benchmarking findings by program features (including behavioral intervention tactics), or (b) screen out specific programs within a program class that do not sufficiently leverage behavioral intervention tactics drawn from the social sciences.

The first level of the taxonomy is the sector – Residential and Commercial. In some cases it was necessary to differentiate benchmarking results by specific audiences or targets in each sector; this will be possible by tagging each program by a number of features (described below), including audience.

The second level of the taxonomy is the family, representing broad-based groups of efforts organized around their primary method of driving action. These families were first proposed for classification of behavior-based energy efficiency efforts in the "ACEEE Field Guide to Utility-Run Behavior Programs."[33] We provide a summary of these families below; for a more complete definition and examples please refer to the Field Guide.

---

[32] *Feedback that does not occur in real-time (as distinct from real-time energy information)*

[33] *Mazur-Stommen, Susan, and Kate Farley. ACEEE Field Guide to Utility-Run Behavior Programs. American Council for an Energy-Efficient Economy, Report Number B132. December 2013.*

- The first family, Cognition, includes efforts that appeal to emotions and/or rely primarily on delivering information to customers as a means of driving change. Traditional, rebate-based programs that communicate program benefits and general savings potential for the program can be considered Cognition efforts. However, for classification as a behavior-based initiative, this information must be delivered using social science-based methods and/or messaging (explained further in the Appendix). The type of information delivered through initiatives in the Cognition family will differ from that delivered in the Calculus family, in that the campaigns falling under Cognition may deliver general or non-customized information.
- In contrast, efforts in the Calculus family typically provide the inputs customers need to make economically rational decisions about saving energy, and may provide customer-specific inputs to enhance effectiveness. Customer-specific feedback, diagnostic efforts like audits, and targeted incentives fall into this category. Again, for classification in this benchmarking effort, we include only those calculus-based efforts that leverage non-price-based behavioral intervention strategies to drive change.
- Efforts that leverage Social Interactions to share information and stimulate change comprise a third family. These interactions could occur between peers, either online or in-person, or through "trusted community members" such as may occur through community-based social marketing or an energy coordinator in a large organization. These interactions could occur between individual users ("human scale") or between organizations.

The third level of the taxonomy is categories of behavioral intervention strategies that fall within each family. These categories designate further differentiation of efforts by the primary delivery mode of the general behavior change approach. Under Cognition, information may be delivered directly to end-users (through education), or through market actors. Under Calculus, quantifiable information required for customers to make investment decisions may be delivered through diagnostic efforts such as home audits that seek to quantify the potential to save, or through feedback efforts that intervene at varying frequencies with information on how much energy customers are using (but may not necessarily provide the synthesis of diagnostic efforts). Under social interactions, we differentiate those interactions that are primarily for information sharing and diffusion (e.g., "human scale" in the residential sector), compared with interactions with some element of competition, such as games.

The fourth level of the taxonomy, and the level around which we structured benchmarking and reporting, is comprised of exemplars of each category, which we interpreted as program classes. The members of this level are grouped by program types as they are typically implemented and evaluated. We describe each program class below the sector-specific taxonomies.[34]

The fifth level of the taxonomy is program features. There are multiple feature classes that we tracked throughout the benchmarking process, to allow greater differentiation of results (if warranted). We sought to identify places where these features matter and highlight results by feature, though we will not attempt to comprehensively address or compare all features.[35] Simply tagging programs in our benchmarking exercise by feature will enable deeper reporting within (or across) program classes where we may see wide variation in results depending on these features.

---

[34] *This list contains working definitions of these program classes. The specific definitions may be refined during the literature review process. Final definitions will be included in the final deliverable.*

[35] *In many cases it may not be possible to differentiate results by a feature such as "behavioral intervention tactic" because programs may layer multiple tactics. However, if results are markedly different among, for example, community-based programs with a competition tactic and those without that tactic, we may highlight those differences that may be associated with including competition.*

For example, a program feature like opt-out engagement or a market transformation objective may be a "feature" in multiple program classes that cut across the taxonomy.

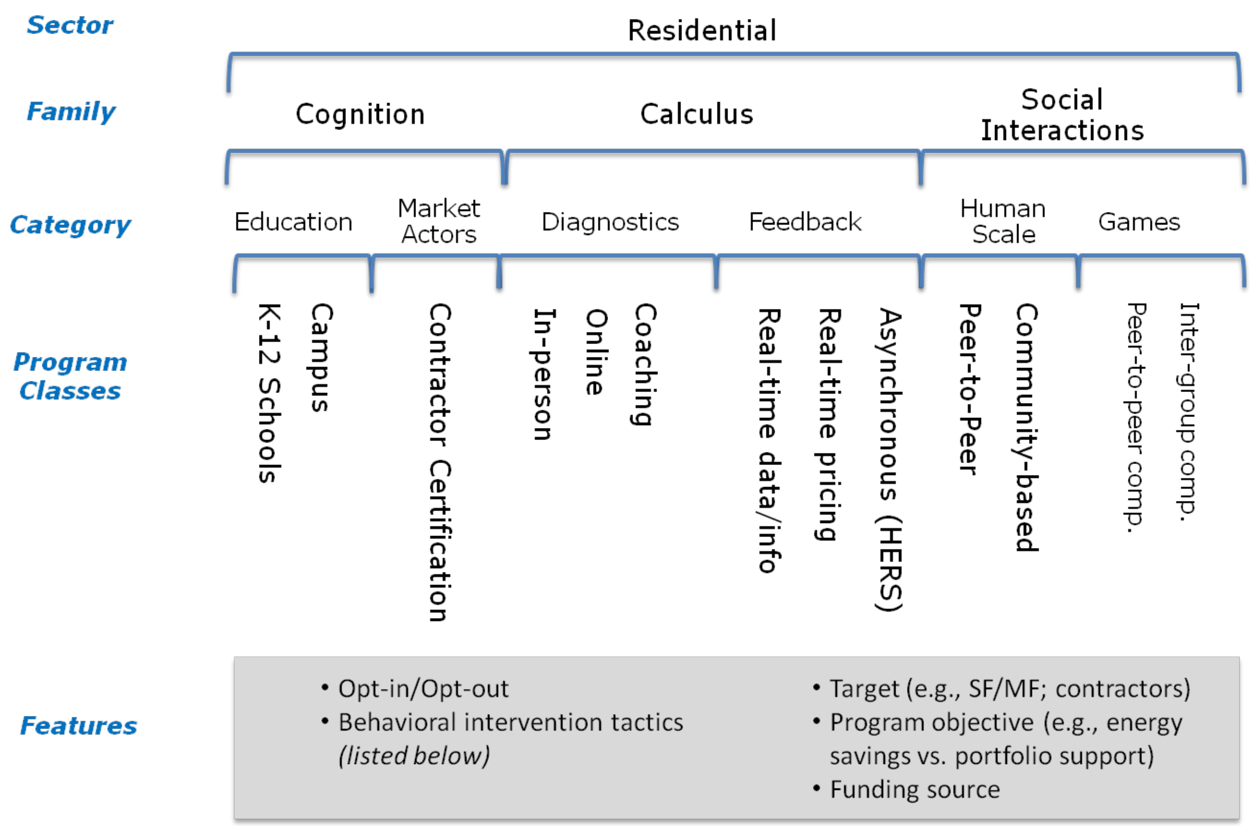
The customer engagement option – an opt-in vs. opt-out program - is one feature that may lead to variation in savings results. We also tracked the target audience and funding source in the event that these features are associated with meaningful differences in outcomes. To acknowledge that behavioral programs may play varying roles in program portfolios, such as generating "direct" savings versus market transformation or promoting other rebate programs, we tagged programs by intended outcome or objective.

Finally, we considered behavioral intervention tactics to be a critical feature worth tracking for this exercise. Behavioral intervention tactics are the various social science-based methods that programs use to stimulate change, such as social norms or in-person interactions. Many programs within a family or program class use multiple tactics, with greater or lesser emphasis, and some tactics are used across multiple program classes. Understanding which combination of tactics led to greater impacts within a fairly similar program design will be critical for Minnesota stakeholders to make programmatic decisions in the future.

Our taxonomy of social science-based and evaluable residential and commercial behavioral programs is below. At the level of "program classes", this graphic represents the programs that we know to be in existence today that meet the criteria above – specifically, meeting the definition of a behavioral program, and having evaluable energy savings. As we conducted our literature review, we found that there were program classes that fell into this taxonomy in theory, but have not yet been implemented or do not have evaluable savings at this point (e.g., other Cognition-based efforts). Because this taxonomy is designed to guide our benchmarking effort, we have not attempted to represent these programs in the taxonomies.[36] However, we believe that the framework we represent – particularly the families of behavioral efforts – is general and comprehensive enough to accommodate future program designs.

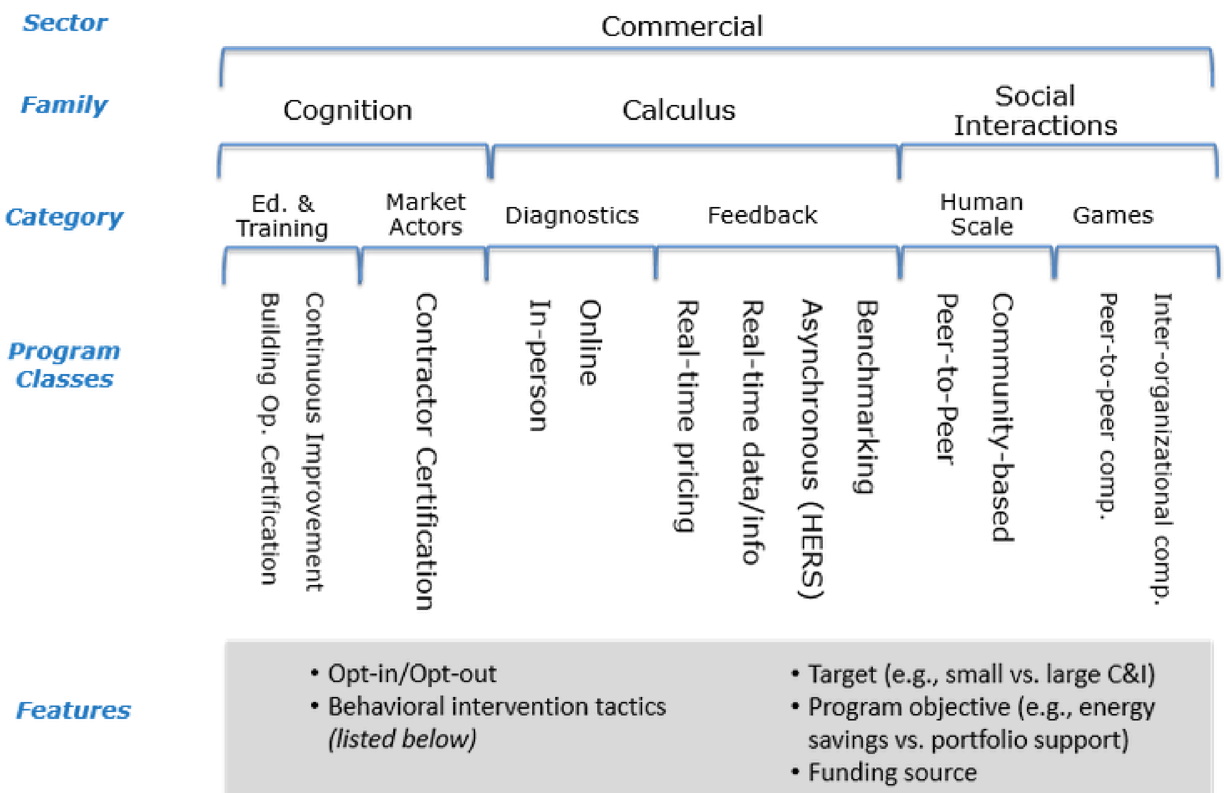
[36] *In other words, this taxonomy represents a subset of the taxonomy in the "ACEEE Field Guide to Utility-Run Behavior Programs", refined based on what we see as determinable and evaluable programs in existence today.*

Preliminary Residential Behavioral Program Taxonomy



The commercial taxonomy shares many elements with the residential taxonomy. Key differences include education and training programs oriented toward facility and building operators; benchmarking programs as distinct from asynchronous feedback programs; and social interactions in which the actors may be organizations rather than individuals (e.g., inter-organizational competitions).

Preliminary Commercial Behavioral Program Taxonomy



The taxonomy was submitted to the Minnesota DER on January 27, 2015, in the same memorandum that detailed our definition for behavioral programs. After discussion with stakeholders in the Minnesota DER, we refined the taxonomy (see figures in section 5). It should be noted that, as we conducted our literature review, some program classes either had very little savings data available or we did not identify many programs that specifically used behavioral strategies within these classes. Other program classes, for example, asynchronous feedback for the residential sector, had significant savings data available, and there were also numerous programs that specifically used behavioral strategies to influence consumer behavior.

While we specifically defined program classes to be included in this study, we also defined types of programs that would be excluded. These program types were excluded because they either lack sufficient savings-related metrics, because they do not specifically use behavior change strategies, or because they do not attempt to change behavior. These included the following:

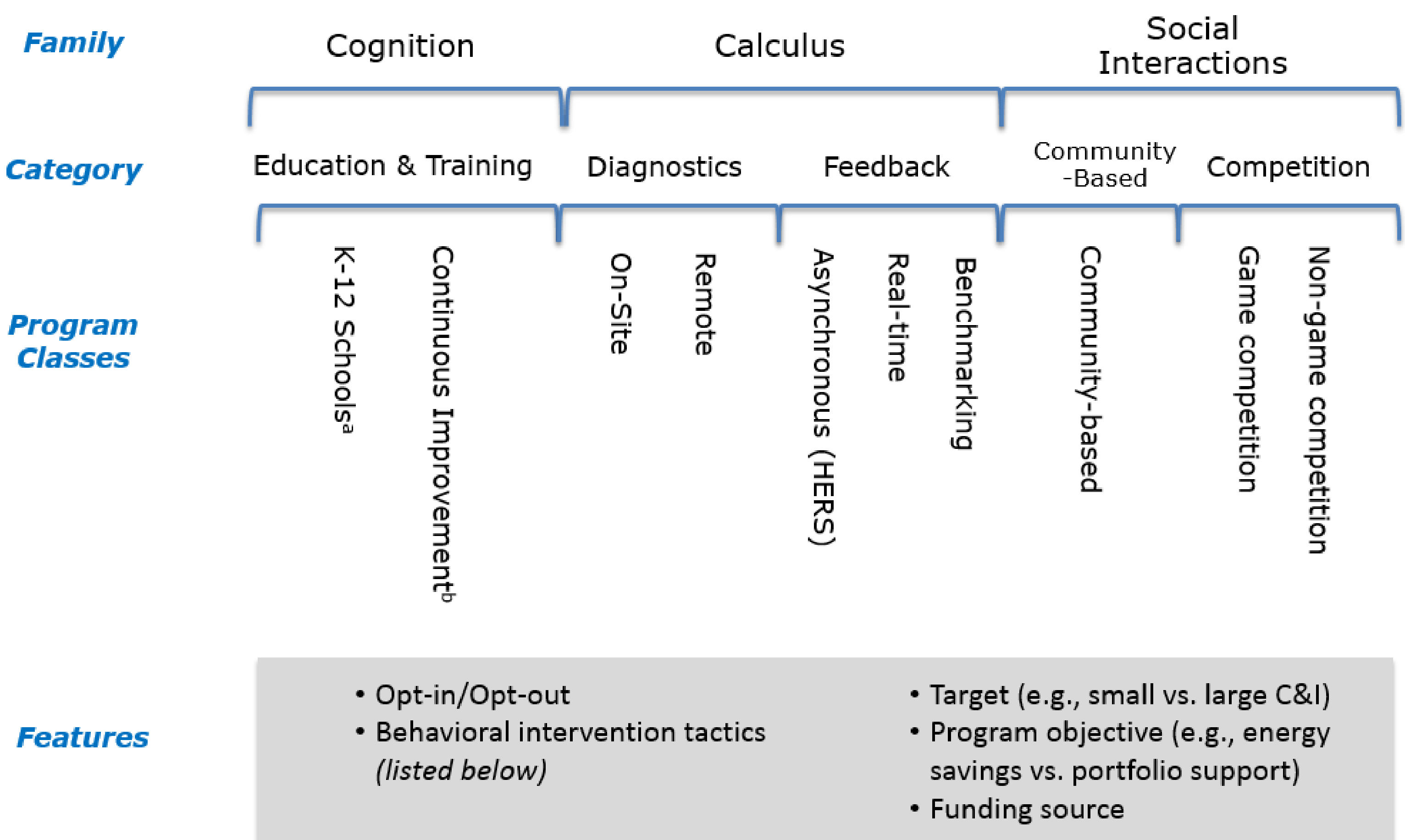
1. "Cognition" -based programs that use media-specific outreach
   a. Mass media outreach.
   b. Social media campaigns.
   c. Providing information on customer bills (information provided on bill may be similar to that provided on a Home Energy Report, but the distinguishing factor elevating a Home Energy Report to a behavior-based program for this study is its reliance on other social science insights, such as framing or peer comparisons, to effect change).
2. Direct install programs
   a. Does not require active behavior change among residents.

3. Home energy audit programs
    a. Programs that promote or encourage conservation behaviors or changes in operating practices (e.g., thermostat set points) alongside measure-based actions in the audit report recommendations. If these recommendations are delivered with traditional tactics that do not employ strategies described above, they were excluded.

However, as we conducted our literature review, we found the below (Figure 11), to be more appropriate. For example, an original classification in our taxonomy included both community-based and peer-to-peer programs. However, because peer-to-peer engagement strategies are common in community-based programs, we do not report on peer-to-peer as stand-alone programs.

We also collapsed the commercial and residential taxonomies to reduce overlap to produce the final taxonomy below.

Figure 12. Final Program Taxonomy for the Residential and Commercial Sectors



[a] *We consider K-12 Schools programs to be a "Commercial" offering based on where the program is delivered, though savings may be achieved in the school or student homes*

[b] *Continuous Improvement (also known as Strategic Energy Management) is a commercial-only offering*

Literature Review

During January and February, 2015, we conducted a literature review of behavioral programs that had been implemented during 2010 or thereafter. This review included research across multiple sources, including databases such as DSM Insights and CALMAC, published papers, peer-reviewed white papers, third-party evaluations, implementer-led studies, and conference proceedings, including those available from ACEEE and IEPEC. As we identified new programs, we first ensured that they met our inclusion criteria. Programs that did meet the study inclusion criteria were then classified per the program taxonomy. In addition to identifying individual behavioral programs, we also reviewed meta-analyses of behavioral programs. In the Findings section of this report, we include relevant information (for example, savings ranges) captured in these meta-analyses. We also used the meta-analyses to ensure that we had captured programs included in these. It is important to note that we did not exhaustively capture all programs that fall within each program class. Instead, we specifically searched for programs that met our inclusion criteria. As such, for some program classes, findings are limited.

In total, we reviewed 170 studies, evaluations, and meta-analyses of behavioral programs. We included a total of 58 programs from those studies in the benchmarking analysis.

Expert Interviews

In order to gather additional evaluation reports relevant to this study, we conducted 20 in-depth interviews with program administrators and evaluators of behavioral feedback programs. We initially contacted those individuals via email to inquire about particular programs (or program classes) and whether evaluations of those programs had been completed. We then conducted telephone interviews. Often, the individuals who we interviewed provided us with additional evaluation reports, and where appropriate, we included those in this study.

Benchmarking Analysis

In order to conduct the benchmarking analysis, we developed several metrics which we used to benchmark behavioral programs against one another. It is important to note that the documentation of behavioral programs often only captures a selection of these metrics. As such, metrics that were publicly available were included. Some behavioral programs may have several metrics that can be recorded, while other programs may have few metrics that can be recorded. In addition, for some program classes, there was insufficient data to conduct benchmarking analyses, primarily due to a lack of comparability between metrics. When this occurred, we report on these program classes to the extent possible, but do not attempt to compare programs to one another.

The primary metric that we used to compare energy savings results across programs and program classes was first-year energy savings[37], expressed as average percent savings per premise. All savings metrics reported in the Benchmarking Analysis Findings section of this report are for savings as a percent of baseline consumption, unless otherwise noted. While we report several point estimates in the Findings section of this report, it is important to note that these may not be

[37] *When available, we tracked these metrics for electricity, gas, and other fuel savings.*

statistically significant (typically due to low sample size). However, these point estimates are still the best available estimate of how a program performed. In addition, in many cases, one might argue that the participant population is not a "sample" (e.g., billing analysis is typically performed with the entire participation population). Thus, for the purpose of this report, we do not typically report sampling error associated with point estimates.

We also collected data on other success metrics for programs, such as participation rate, and incremental savings. Where available, we tracked metrics over time for long-term programs with multiple years of program data. We had also planned to search for cost-effectiveness metrics. However, this information was not readily available for the majority of program evaluations that we reviewed. Most frequently, this information is reported at the portfolio (and not program) level, limiting its utility to this study. In the few instances in which were able to identify relevant cost-effectiveness metrics, we do report on it.

The below table includes a list of metrics that we recorded for behavioral programs included in our study. As noted above, not all programs had publicly available data related to all metrics. As such, we catalogued metrics that were available for the programs included in our study.

Table 18. List of Metrics with Abbreviated Definitions

Metric Class	Metrics Recorded (when available)
Taxonomy Membership	• Sector • Family • Category • Class
Program Features (as represented in taxonomy)	• Opt-in/opt-out • Behavioral intervention tactics • Target of intervention • Program objectives that include energy savings • Funding sources
Categorical Program Characteristics	• Electric or gas • Program administrator • Program implementer • State(s) where program was implemented • Evaluation design • Program targeting criteria • Program participant characteristics • Report/feedback frequency for feedback programs
Numeric Program Characteristics	• Program year start date • Number of participating customers • Opt-in rate/opt-out rate
Energy Savings Demand Savings	• First year energy % savings per premise • First year energy unit savings (kWh, therm)
Subsequent Year Energy Savings	• Treatment continued or discontinued • Energy % savings per premise • Energy unit savings (kWh, therm) per premise
Demand Savings Metrics	• Summer peak load reduction (%) • Winter peak load reduction (%)

Metric Class	Metrics Recorded (when available)
Cross-Participation	• Incremental % of behavioral program participants participating in other energy efficiency programs • Incremental savings per premise (sometimes known as "double-counted" savings*)[38]
Actions Taken	• Actions targeted by behavioral intervention • Actions taken in response to program intervention

Reporting

During January and February, 2015, we submitted memoranda to the Minnesota DER. These memorandum described the definition of behavioral programs to be used in this study, development of the program taxonomy, literature review objectives, and metrics to be used for the benchmarking analysis. All memoranda were reviewed by Minnesota DER stakeholders.

[38] *The savings associated with increased uptake in other energy efficiency programs that is attributable to the behavioral program, but sometimes considered "double-counted" savings, because the other rebate/DSM programs typically claim these savings.*

References

Abrahamse W., Steg L., Vlek C., & Rothentagger T. (2005). A review of intervention studies aimed at household energy use. *Journal of Environmental Psychology*, 25:273-291.

Alberta Energy Efficiency Alliance. (2014). *Advancing Energy Efficiency in Calgary: Energy Savings Through Consumer Feedback Programs*. Available at: http://www.aeea.ca/pdf/calgary-advancing-energy-efficiency.pdf

Alcott H. (2011). Social norms and energy conservation. *Journal of Public Economics*, 95.

Alcott H. and Mullainathan. (2010). Behavior and energy policy. *Science*, 327(5970):1204-1205.

Alliance to Save Energy. (2012). *Energy Saving Success Stories from Southern California 2011-2012*. Retrieved from http://www.ase.org/sites/ase.org/files/resources/Media%20browser/powersave_schools-success_book_2011-12.pdf

Ashby K., Forster H., Ceniceros B., Wilhelm B., Friebel K., Henschel R., and Samiullah S. (2012). *Green with envy: neighbor comparisons and social norms in five home energy report programs*. ACEEE Summer Study on Energy Efficiency in Buildings. Retrieved from http://www.aceee.org/files/proceedings/2012/data/papers/0193-000218.pdf

BC Hydro. (2014). *Demand Side Management Milestone Evaluation Summary Report F2013*. Retrieved from http://www.bchydro.com/content/dam/BCHydro/customer-portal/documents/corporate/regulatory-planning-documents/revenue-requirements/directive-66-f2013-demand-side-management-milestone-evaluation-summary-report.pdf

BC Hydro. (2015). *Demand Side Management Milestone Evaluation Summary Report F2014*. Retrieved from http://www.bchydro.com/content/dam/BCHydro/customer-portal/documents/corporate/regulatory-planning-documents/revenue-requirements/2015-01-14-bch-d66-f2014-rpt.pdf

Bensch, I. (2013). *Identifying the impacts of Cool Choices' game at Miron Construction: Energy savings from player actions*. Retrieved from http://www.ecw.org/publications/identifying-impacts-cool-choices-game-miron-construction-energy-savings-player-actions

Bernedo, M., Ferraro, P., & Price, M.. (2014). The Persistent Impacts of Norm-based Messaging and their Implications for Water Conservation. *Journal of Consumer Policy*, *1*(16). Retrieved from http://www.kysq.org/docs/Bernedo.pdf

Braithwait S., Hansen D., and Hillbrink M. (2013). *2012 load impact evaluation of San Diego Gas and Electric's peak time rebate program.* Christensen Associates Energy Consulting. Inc. CALMAC Study ID: SDG0266.01.

Buchanan K., Russo R., & Anderson B. (2015). The question of energy reduction: The problem(s) with feedback. *Energy Policy*, *77*(0), 89–96. http://doi.org/10.1016/j.enpol.2014.12.008

Cadmus. (2009). *FY2009 Energy Efficiency Portfolio Evaluation Summary*. Retrieved from http://www.srpnet.com/about/pdfx/Summary.pdf

Cadmus. (2013a). *Impact Evaluability Assessment of California's Continuous Energy Improvement Pilot Program.* Retrieved from http://www.calmac.org/publications/CA_CEI_Pilot_Evaluability_Assessment_REVISED_FINAL.pdf

Cadmus. (2013b). *Energy Management Pilot Impact Evaluation: A Report to the Bonneville Power Administration*. Retrieved from http://www.bpa.gov/EE/Utility/research-archive/Documents/BPA_Energy_Management_Impact_Evaluation_Final_Report_with_Cover.pdf

Cappers P., Todd A., Perry M., Neenan B., and Boisvert R. (2013). *Quantifying the Impacts of Time-based Rates and Other Treatments in Consumer Behavior Studies: Protocols and Guidelines.* Lawrence Berkeley National Laboratory. Retrieved from http://emp.lbl.gov/sites/all/files/lbnl-6301e.pdf

CEE. (2012). *Summary of Commercial Whole Building Performance Programs: Continuous Energy Improvement and Energy Management and Information Systems.* Retrieved from http://library.cee1.org/content/summary-commercial-whole-building-performance-programs-continuous-energy-improvement-and-en-1

Center for Energy and Environment. (2014). *Verification of Savings from Xcel Energy Minnesota's Print Energy Feedback Pilot Project - Final Report*.

Davis, M. (2011). *Behavior and Energy Savings Evidence from a Series of Experimental Interventions*. Retrieved from http://blogs.edf.org/energyexchange/files/2011/05/BehaviorAndEnergySavings.pdf

Delmas M., Fischlein M., and Asensio O. (2013). Information strategies and energy conservation behavior: a meta-analysis of experimental studies from 1975-2011. *Energy Policy* 61, 729-739.

Dethman, L., Stewart, J., & Thomsen, K. (2013). *Integrating Process and Impact Findings to Understand and Measure Behavioral Savings at Work.* Presented at the International Energy Program Evaluation Conference, Chicago, IL. Retrieved from http://www.cadmusgroup.com/wp-content/uploads/2014/08/Cadmus_IEPEC2013_BehavioralSavingsWork.pdf

Distributed Energy Financial Group LLC (2014). *How Do Prepay Electricity Programs Impact Consumer Behavior.* Retrieved from http://defgllc.com/publication/how-do-prepay-electricity-programs-impact-consumer-behavior/

DNV-GL. (2014a). *2013 Home Energy Reports Evaluation*. Prepared for Puget Sound Energy. Retrieved from http://opower.com/company/library/verification-reports

DNV-GL. (2014b). *Evaluating Opt-In Behavior Programs: Issues, Challenges, and Recommendations*. Prepared for California Public Utilities Commission. Retrieved from http://www.calmac.org/%5C%5C//publications/DNV-GLEvaluationofOpt-InPrograms_Final_V7_9-11-2014.pdf

DNV KEMA Energy & Sustainability. (2012a). *Puget Sound Energy's Home Energy Reports Program Three Year Impact, Behavioral and Process Evaluation*. Prepared for Puget Sound Energy. Retrieved from http://opower.com/company/library/verification-reports

DNV KEMA Energy & Sustainability. (2012b). *Puget Sound Energy's Home Energy Reports Program: Three Year Impact, Behavioral and Process Evaluation* (Impact, Behavioral, and Process Evaluation). Retrieved from http://opower.com/company/library/verification-reports

DNV KEMA, Inc. (2013). *Review of PG&E Home Energy Reports Initiative Evaluation* (Joint Savings Impact Analysis). Retrieved from http://opower.com/company/library/verification-reports

DNV KEMA, & Research into Action. (2014). *NEEA Industrial Initiatives - Market Progress Evaluation Report #8*. Retrieved from https://neea.org/docs/default-source/reports/neea-industrial-initiatives--market-progress-evaluation-report-8.pdf?sfvrsn=10

Donnelly, K. (2014). *Connecticut Neighbor to Neighbor Energy Challenge: Better Buildings Neighborhood Program Final Report*. Retrieved from http://www.osti.gov/scitech/servlets/purl/1114148

Dougherty, A. (2013). *Massachusetts cross-cutting behavioral program evaluation: integrated report.* Opinion Dynamics, Waltham, MA. Retrieved from http://www.rieermc.ri.gov/documents/2013%20Evaluation%20Studies/ODC_2013_Cross_Cutting_Behavioral_Program_Evaluation.pdf

Dougherty, A., & Provencher, B. (2013, August). *Impacts of Feedback Programs: Generating Comparable Impacts across Varying Program Design Models*. Presented at the International Energy Program Evaluation Conference, Chicago, IL.

DTE Energy. (2014a). *DTE Energy Smart Currents Dynamic Peak Pricing Pilot: Final Report*. Retrieved from https://www.smartgrid.gov/sites/default/files/doc/files/DTE-SmartCurrents_FINAL_Report_08152014.pdf

DTE Energy. (2014b). *Pre-Pay Pilot (Pay As You Go) Program – 2014 Annual Report.* Retrieved from https://efile.mpsc.state.mi.us/efile/docs/16457/0007.pdf

Duke Energy. (2014). *Duke Energy Progress: Prepay Pilot Program Final Report.* Retrieved from http://dms.psc.sc.gov/pdf/matters/61E8CAC6-155D-141F-235A497540708F60.pdf

Edison Foundation. (2013). Innovations Across the Grid: 2013 IEE Partner Roundtable.

Ehrhardt-Martinez, K., Donnelly, K., & Laitner, S. (2010). *Advanced Metering Initiatives and Residential Feedback Programs: A Meta-Review for Household Electricity-Saving Opportunities*. Retrieved from http://www.energycollection.us/Energy-Metering/Advanced-Metering-Initiatives.pdf

Electric Power Research Institute. (2010). *Paying Upfront: A Review of Salt River Project's MPower Prepaid Program*. Retrieved from http://www.srpnet.com/environment/earthwise/pdfx/spp/EPRI_MPower.pdf

Energy Center of Wisconsin. (2014). *Minnesota Power's Advanced Metering Infrastructure Project AMI Behavioral Research Pilot – Phase 1*. Retrieved from https://www.smartgrid.gov/sites/default/files/doc/files/MN%20Power%20CBP%20interim%20report%20FINAL%20with%20March6%20date.pdf

Energy Trust of Oregon, & The Cadmus Group. (2014). *2014 Energy Trust Workshops on Strategic Energy Management Impact Evaluation*. Retrieved from http://energytrust.org/library/reports/SEM_Evaluation_Workshop_Report.pdf

Erickson, K. (2012). *Summary of Commercial Whole Building Performance Programs Continuous Energy Improvement and Energy Management and Information Systems*. Boston, MA: Consortium for Energy Efficiency. Retrieved from http://library.cee1.org/sites/default/files/library/9235/CEE_CommBldg_WBCEIEMISProgSummPublicVersion_1May2012_0.pdf

Faruqui, A., & Sergici, S. (2010). *Household Response to Dynamic Pricing of Electricity - A Survey of the Empirical Evidence*. Retrieved from http://ssrn.com/abstract=1134132

Faruqui, A. (2013). *Dynamic Pricing: The Bridge to a Smart Energy Future*. Presented at the World Smart Grid Forum, Berlin, Germany. Retrieved from http://www.brattle.com/system/publications/pdfs/000/004/925/original/Dynamic_pricing_-_the_bridge_to_a_smart_energy_future_Faruqui_World_Smart_Grid_Forum_092513.pdf?1380118695

Freeman, Sullivan & Co. (2013). *Evaluation of Pacific Gas and Electric Company's Home Energy Report Initiative for the 2010–2012 Program*. Retrieved from http://www.calmac.org/publications/2012_PGE_OPOWER_Home_Energy_Reports__4-25-2013_CALMAC_ID_PGE0329.01.pdf

Fuller M., Kunkel C., Zimring M., Hoffman I., Soroye K., and Goldman C. (2010). *Driving demand for home energy improvements.* Lawrence Berkeley National Laboratory. Retrieved from http://drivingdemand.lbl.gov/

Gaynoll C., & Patrice Ignelzi. (2013). New Kid in Town: SEM Joins the Demand Management Family. Presented at the Association of Energy Service Professionals Natoinal Conference, Orlando, FL. Retrieved from http://library.aesp.org/resources/DocuWorks/file_display.cfm?id=1262

Glinsmann, B., & Provencher, B. (2013, August). *I can't use a Randomized Controlled Trial – Now What? Comparison of Methods for Assessing Impacts from Opt-In Behavioral Programs.* Presented at the International Energy Program Evaluation Conference (IEPEC), Chicago, IL.

Grossberg, F., Wolfson, M., Mazur-Stommen, S., Farley, K., & Nadel, S. (2015). *Gamified Energy Efficiency Programs* (No. B1501). American Council for an Energy Efficient Economy. Retrieved from http://www.aceee.org/research-report/b1501

Hall, N., & Roth, J.. (2014). *Impact Evaluation of the Smart Energy Now Program (NC) (Pilot)*. Retrieved from http://starw1.ncuc.net/NCUC/ViewFile.aspx?Id=d9f92456-1403-4809-8f35-a31ad2453ea3

Hall, N., Roth, J., & Best, C. (2006). *California energy efficiency evaluation protocols: technical, methodological and reporting requirements for evaluation professionals*. Retrieved from http://www.calmac.org/events/evaluatorsprotocols_final_adoptedviaruling_06-19-2006.pdf

Hall, N., Ladd, D., Roth, J., & Yin, C. (2013). *Process Evaluation of the Smart Energy Now Pilot Program Implemented in Charlotte, NC as a component of Envision Charlotte Sustainability Initiative*. Prepared for Duke Energy.

Hampton J., Hummer J., & Wobus N. (2012*). EnergySmart program: progress review*. Prepared for Boulder County by Navigant, Inc. Retrieved from: http://energysmartyes.com/files/EnergySmart_Navigant_ProgressReport_2012_FINAL.pdf

Harding, M., & Hsiaw A. (2014). Goal setting and energy conservation. *Journal of Economic Behavior & Organization*, 107, Part A, 209–227.

Ho, D.E., Kosuke I., Gary K., and Elizabeth S. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis* 15(3): 199-236.

Ignelzi, P.J., Randazzo K., Dethman L., and Lutzenhiser L. (2013). *Paving the way for a richer mix of residential behavior programs.* Retrieved from: http://www.calmac.org/publications/residential_behavior_white_paper_5-31-13_final.pdf

Illume Advising. (2014). *MyMeter Multi-Utility Impact Findings*. Retrieved from http://www.midwestenergynews.com/wp-content/uploads/2014/07/Illume_Case_Study.pdf

Illume Advising, & Navigant Consulting. (2014). *Rhode Island Behavioral Program and Pilots Impact and Process Evaluation*. Retrieved from http://www.rieermc.ri.gov/documents/2014%20Evaluation%20Studies/Rhode%20Island%20Behavioral%20Program%20and%20Pilots%20Impact%20and%20Process%20Evaluation.pdf

IndEco Strategic Consulting. (2011). *In-Depth Review of Energy Efficiency Programs*. Toronto, ON, Canada. Retrieved from http://www.exec.gov.nl.ca/exec/ccee/publications/canadian_energy_efficiency_programs_part_b.pdf

Integral Analytics, Inc. (2012). *Impact & Persistence Evaluation Report Sacramento Municipal Utility District Home Energy Report Program*. Retrieved from http://opower.com/company/library/verification-reports

Ito, K., Ida, T., & Tanaka, M. (2015). *The Persistence of Moral Suasion and Economic Incentives: Field Experimental Evidence from Energy Demand*. Retrieved from https://www.aeaweb.org/aea/2015conference/program/retrieve.php?pdfid=940

Itron. (2013). *2010-2012 CPUC HEES Impact Evaluation*. Prepared for: California Public Utilities Commission. Retrieved from http://www.calmac.org/publications/HEES_Final_Report_20130708.pdf

Johnson, P., Xu, Y., Brewer, R., Lee, G., Katchuck, M., and Moore C. (2013). *Beyond kWh: Myths and Fixes for Energy Competition Game Design.* Presentation to Energy 2013 conference, Lisbon. March. http://meaningfulplay.msu.edu/proceedings2012/mp2012_submission_72.pdf. https://csdl-techreports.googlecode.com/svn/trunk/techreports/2012/12-12/12-12.pdf.

KEMA. (2010). *Puget Sound Energy's Home Energy Reports Program: 20 Month Impact Evaluation*. Retrieved from http://opower.com/company/library/verification-reports

KEMA, Inc. (2013). *Puget Sound Energy's Home Energy Reports 2012 Impact Evaluation*. Retrieved from http://opower.com/company/library/verification-reports

Khawaja, M., and Stewart, J. (2014/2015). *Long-run savings and cost-effectiveness of home energy report programs.* The Cadmus Group, Inc.

Kiselwich, R. (2014). *Successful Rollout of Dynamic Pricing: BGE's Smart Energy Rewards*. Presented at the National Town Meeting on Demand Response and Smart Grid, Washington, DC. Retrieved from http://www.demandresponsetownmeeting.com/wp-content/uploads/2012/03/Kiselewich-Ruth-5-DR.pdf

Kraus, Y. (2014). *Bainbridge Energy Challenge. Energy efficiency and conservation block grant (EECBG) - Better buildings neighborhood program. Final Technical Report.* (No. DOE-BBNP--0003805). US Department of Energy. Retrieved from http://www.osti.gov/scitech/servlets/purl/1120149

Mazur-Stommen, S., and Foster, B. (2012). *Results from Recent Real-Time Feedback Studies*. Retrieved from http://www.aceee.org/sites/default/files/publications/researchreports/b122.pdf

Mazur-Stommen, S. and Farley, K. (2013). *ACEEE field guide to utility-run behavior programs*. Retrieved from http://www.aceee.org/research-report/b132

Megdal & Associates, & Albert, S. (2012). *Energy $mart Focus Benchmarking of Institutions and Commercial Real Estate Components Program Impact Evaluation Report*. Prepared for The New York State Energy Research and Development Authority. Retrieved from http://www.nyserda.ny.gov/About/Program-Evaluation/Evaluation-Contractor-Reports/2012-Reports

Miknaitis G., Lux J. Dyanko D, Hamann M, & Burns W. (2014). *Tapping Energy Savings from an Overlooked Source: Results from Behavioral Change Pilot Program Targeting Mid-Sized Commercial Customers*. Retrieved from https://www.aceee.org/files/proceedings/2014/data/papers/7-153.pdf

Minnesota Department of Energy Resources. (2012). *Analysis, Recommendations, and Proposed Decision*. Retrieved from https://www.edockets.state.mn.us/EFiling/edockets/searchDocuments.do?method=showPoup&documentId={AB447C13-1195-48CA-8239-E732471C62E1}&documentTitle=20122-70946-01

Minnesota Statutes. Energy Conservation Improvement, 216B.241 (2014). Retrieved from https://www.revisor.mn.gov/statutes/?id=216B.241

Navigant Consulting. (2010). *Energy Efficiency/ Demand Response Plan: Plan Year 2 Evaluation Report: OPOWER Pilot*. Prepared for Commonwealth Edison. Retrieved from http://opower.com/company/library/verification-reports

Navigant Consulting. (2011). *Evaluation Report: OPOWER SMUD Pilot Year 2*. Retrieved from http://opower.com/company/library/verification-reports

Navigant Consulting. (2012a). *Energy Efficiency / Demand Response Plan: Plan Year 3 Evaluation Report: Home Energy Reports*. Prepared for Commonwealth Edison. Retrieved from http://opower.com/company/library/verification-reports

Navigant Consulting. (2012b). *EnergySmart Program Progress Review*. Prepared for Boulder County. Retrieved from http://www.energysmartyes.com/files/EnergySmart_Navigant_ProgressReport_2012_FINAL.pdf

Navigant Consulting. (2014a). *Home Energy Reports Program PY5 Evaluation Report*. Prepared for Commonwealth Edison. Retrieved from http://ilsagfiles.org/SAG_files/Evaluation_Documents/ComEd/ComEd%20EPY5%20Evaluation%20Reports/ComEd_PY5_HER_Report_2014-01-28_Final.pdf

Navigant Consulting. (2014b). *ComEd Commercial & Industrial Behavioral Program Evaluation*. Retrieved from http://ilsagfiles.org/SAG_files/Evaluation_Documents/ComEd/ComEd%20EPY5%20Evaluation%20Reports/ComEd_EPY5_C&I_Behavioral_Program_Eval_Report_2014-03-11_Final.pdf

Navigant Consulting. (2015). *C3-CUB Energy Saver Program PY6 Evaluation Report DRAFT*. Retrieved from http://ilsagfiles.org/SAG_files/Evaluation_Documents/Draft%20Reports%20for%20Comment/PY6_Reports/ComEd/ComEd_PY6_CUB-C3_Evaluation_Report_2015-01-15.pdf

Navigant Consulting, Opinion Dynamics, & Itron. (2012). *Energy Efficiency / Demand Response Plan: Plan Year 4 Evaluation Report: Home Energy Reports*. Prepared for Commonwealth Edison. Retrieved from http://opower.com/company/library/verification-reports

New Jersey Natural Gas. (2014). *New Jersey Natural Gas' Conserve to Preserve Community Rewards Provides Fundraising Opportunity For Schools.* Wall, NJ. Retrieved from http://www.njresources.com/news/releases/2014/njng/14-16Re-launchOfCTPCommunRwrds.asp

Newsham, G. R., and Bowker, B. G. (2010). The effect of utility time-varying pricing and load control strategies on residential summer peak electricity use: A review. *Energy Policy*, 38. Retrieved from http://sedc-coalition.eu/wp-content/uploads/2011/05/Newsham-10-03-26-Dynamic-Pricing-and-Load-Control-Literature-Review.pdf

New York State Customer Engagement Committee. (2014). *Reforming the Energy Vision (REV) Working Group I: Customer Engagement. Staff Report on the Work of the Customer Engagement Committee.* Prepared for New York Department of Public Service. Retrieved from http://www3.dps.ny.gov/W/PSCWeb.nsf/96f0fec0b45a3c6485257688006a701a/853a068321b1d9cb85257d100067b939/$FILE/WG%201_Customer%20Engagement_Final%20Report%20&%20Attachments.pdf

Nexant. (2014a). *Load Impacts and Customer Choice Results from SMUD's Two-Year Smart Pricing Options Pilot*. Presented at the National Town Meeting on Demand Response and Smart Grid

Nexant. (2014b). *HAN Phase 3 Impact and Process Evaluation Report.* Prepared for Pacific Gas and Electric. Retrieved from http://www.calmac.org/publications/HAN_Impacts_and_Savings_Report_FINAL.pdf

Opinion Dynamics. (2011). *Evaluation of National Grid's Community Pilot Program Energy Action: Aquidneck and Jamestown.* Prepared for National Grid. Retrieved from http://www.rieermc.ri.gov/documents/evaluationstudies/2011/Aquidneck%20Pilot%20Evaluation_Final%20Report_10-07-2011.pdf

Opinion Dynamics. (2012). *Impact and Process Evaluation of 2011 (PY4) Ameren Illinois Company Behavioral Modification Program*. Retrieved from http://opower.com/company/library/verification-reports

Opinion Dynamics. (2014a). *Impact and Process Evaluation of Ameren Illinois Company's Behavioral Modification Program (PY5)*. Retrieved from http://opower.com/company/library/verification-reports

Opinion Dynamics. (2014b). *National Grid Rhode Island System Reliability Procurement Pilot: 2012-2013 Focused Energy Efficiency Impact Evaluation*. Prepared for National Grid. Retrieved from http://www.ripuc.org/eventsactions/docket/4528-NGrid-2015-SRP-Plan%2810-31-14%29.pdf

Opinion Dynamics. (2014c). *Rhode Island EmPower Pilot: Process & Impact Evaluation*. Prepared for National Grid.

Opinion Dynamics. (2014d). *Massachusetts Cross-Cutting Evaluation: Home Energy Report Savings Decay Analysis.*

Opower. (2013). *Fitting a Square Peg into a Round Hole: Behavioral Energy Efficiency at Scale.* Retrieved from http://www2.opower.com/l/17572/2013-03-28/7yh2y

Otter Tail Power. (n.d.). *Town Energy Challenge Pilot*. Retrieved from http://www.cleanenergyresourceteams.org/sites/default/files/OtterTail_ChallengePilot.pdf

Ozog, M., & Integral Analytics. (2013). *The Effect of Prepayment on Energy Use*. Commissioned by the DEFG PrePay Energy Working Group. Retrieved from http://www.exceleron.com/wp-content/uploads/2014/01/The-Effect-of-Prepayment-on-Energy-Use.pdf

PWP, Inc, & Michaels Energy. (2014). *Energy Trust of Oregon Strategic Energy Management Intro Pilot Evaluation Report*. Retrieved from http://energytrust.org/library/reports/SEMi_Report_140604.pdf

Reguly. (2013). The role of behavioral programs approaches in utility customer experience. Presented at BECC

Research into Action. (2013). *Preliminary Energy Savings Impact Evaluation: Better Buildings Neighborhood Program*. Prepared for US Department of Energy. Retrieved from http://www1.eere.energy.gov/analysis/pdfs/energy_savings_impact_bbnp_110413.pdf

Sami, K., M., & Stewart, J. (2014). *Long-Run Savings and Cost-Effectiveness of Home Energy Report Programs*. Retrieved from http://www.cadmusgroup.com/wp-content/uploads/2014/11/Cadmus_Home_Energy_Reports_Winter2014.pdf

SBW Consulting. (n.d.). *Puget Sound Energy Resource Conservation Management*. Prepared for Puget Sound Energy.

Schultz P.W., Nolan J.M., Cialdini R.B., Goldstein N.J., and Vladas G. (2007). The constructive, destructive, and reconstructive power of social norms. *Psychological Science*, 18:429-439

Sergici, S. (2011). *Measurement and Verification Principles for Behavior-Based Effiency Programs*. Retrieved from https://opower.com/uploads/library/file/17/opower-brattle_group-mv_principles_document-_final.pdf

Sergici, S., and Faruqui, A. (2011). Measurement and Verification Principles for Behavior-Based Energy Efficiency Programs. Prepared for OPOWER. Retrieved from https://opower.com/uploads/library/file/17/opower-brattle_group-mv_principles_document-_final.pdf

Smith, B. (2014). *http://beccconference.org/wp-content/uploads/2014/12/presentation_Smith_2.pdf*. Presented at the BECC, Washington, DC. Retrieved from http://beccconference.org/wp-content/uploads/2014/12/presentation_Smith_2.pdf

Southern California Edison. (2013). *2013 Edison SmartConnect Demand Response and Energy Conservation Annual Report*. Retrieved from https://www.pge.com/regulation/DemandResponseOIR/Pleadings/SCE/2013/DemandResponseOIR_Plea_SCE_20130501_273713.pdf

State and Local Energy Efficiency Action Network. (2012). *Evaluation, Measurement, and Verification (EM&V) of Residential Behavior-Based Energy Efficiency Programs: Issues and Recommendations*. Retrieved from https://www4.eere.energy.gov/seeaction/system/files/documents/emv_behaviorbased_eeprograms.pdf

Summit Blue Consulting, LLC. (2009). *Impact Evaluation of OPower SMUD Pilot Study*. Retrieved from http://opower.com/company/library/verification-reports

Syring T., and Laube D. (2013). *Emphasizing the best options for saving energy: final report on the single recommendation strategy CARD grant*. Minnesota Department of Commerce, Division of Energy Resources report COMM/OES-04042011-37273. Retrieved from: https://www.cards.commerce.state.mn.us/CARDS/security/search.do?method=showPoup&documentId={AA59D6B1-1467-4044-BDE9-CE240CA90A54}&documentTitle=85668&documentType=6

Syring T., Gorell A., & Laube D. (2013). *Energy management teams: coordinator resource pilot project.* Minnesota Department of Commerce, Division of Energy Resources report COMM/OES-01192010-B43213. Retrieved from: https://www.cards.commerce.state.mn.us/CARDS/security/search.do?method=showPoup&documentId={A6C93533-FEB8-4597-A7E5-15433B1BF065}&documentTitle=82886&documentType=6

TecMarketWorks. (2014). *Smart Energy Now Impact Evaluation*. Prepared for Duke Energy.

Tendril Networks, & Opinion Dynamics. (2014). *TEP Power Partners Project: Final Report*. Retrieved from http://www.osti.gov/scitech/servlets/purl/1123882.

Todd, A. (2013). *Residential Customer Enrollment in Time-based Rate and Enabling Technology Programs.* Retrieved from http://emp.lbl.gov/sites/all/files/lbnl-6247e.pdf

Todd, Stuart, Schiller, & Goldman. (2012). *Evaluation, measurement, and verification for behavior-based energy efficiency programs: issues and recommendations*. ACEE Summer Study on Energy Efficiency in Buildings.

USDOE. (2013). *Analysis of Customer Enrollment Patterns in Time-Based Rate Programs: Initial Results from the SGIG Consumer Behavior Studies*. Retrieved from http://energy.gov/sites/prod/files/2013/07/f2/CustomerEnrollmentPatterns.pdf

USDOE. (2014). *Customer Participation in the Smart Grid - Lessons Learned*. Retrieved from http://energy.gov/sites/prod/files/2014/10/f18/SG-CustParticipation-Sept2014.pdf

US Environmental Protection Agency. (2012). *Energy Star Portfolio Manager: Benchmarking and Energy Savings Data Trends*. Retrieved from http://www.energystar.gov/sites/default/files/buildings/tools/DataTrends_Savings_20121002.pdf

Vaidya, R., Reynolds, A., Azulay, G., Barclay, D., & Tolkin, B. (2009). *ENERGY STAR® Portfolio Manager and Utility Benchmarking Programs: Effectiveness as a Conduit to Utility Energy Efficiency Programs*. Presented at the Energy Program Evaluation Conference, Portland, OR. Retrieved from http://energy.gov/sites/prod/files/2014/05/f15/1886526335.pdf

Vine, E., and Jones, C. (2015). *A review of energy reduction competitions studies: scaling up deeper savings through comparative feedback and recognition*. Prepared for California Public Utilities Commission. California Institute for Energy and Environment.

Vine, E., Sullivan, M., Lutzenhiser, L., Blumstein, C., & Miller, B.. (2014). Experimentation and the Evaluation of Energy Efficiency Programs. *Energy Efficiency*, 7(4), 627–640.

Vine, Delsey, Buys, L., & Morris, Peter. (2013). *The Effectiveness of Energy Feedback for Conservation and Peak Demand: A literature review*. Retrieved from 10.4236/ojee.2013.21002

Volkman, J., Schick, S., Kesting, O., & Belkhayat, K. (2014). *Energy Trust of Oregon and Commercial Strategic Energy Management: A Catalyst for Accelerating Customer Energy Savings*. Presented at the ACEEE. Retrieved from http://aceee.org/files/proceedings/2014/data/papers/4-616.pdf